

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**Integrating host population contact structure
and pathogen whole-genome sequence data to
understand the epidemiology of infectious diseases**

A thesis presented in partial fulfilment of the requirements for the
degree of Doctor of Philosophy

Massey University, Manawatū
New Zealand

Sabrina S. Greening

2020

Abstract

With advances in high-throughput sequencing technologies, computational biology, and evolutionary modelling, pathogen sequence data is increasingly being used to inform infectious disease outbreak investigations; supporting inferences on the timing and directionality of transmission as well as providing insights into pathogen evolutionary dynamics and the development of antimicrobial resistance. This thesis focuses on the application of pathogen whole-genome sequence data in conjunction with social network analysis to investigate the transmission dynamics of two important pathogens; *Campylobacter jejuni* and *Staphylococcus aureus*.

The first four studies centre around the recent emergence of an antimicrobial resistant *C. jejuni* strain that was found to have rapidly spread throughout the New Zealand commercial poultry industry. All four studies build on the results of an industry survey that were not only used to determine the basic farm demographics and biosecurity practices of all poultry producers, but also to construct five contact networks representing the on- and off-farm movement patterns of goods and services. Contact networks were used in study one to investigate the relationship between farm-level contact risk pathways and the reported level of biosecurity. However, despite many farms having a number of contact risk pathways, no relationship was found due to the high level of variability in biosecurity practices between producers.

In study two the contact risk between commercial poultry, backyard poultry and wild birds was investigated by examining the spatial overlap between the commercial contact networks and (i) all poultry transactions made through the online auction website

TradeMe® and, (ii) all wild bird observations made through the online citizen science bird monitoring project, eBird, with study results suggesting that the greatest risk is due to the growing number of online trades made over increasingly long distances and shorter timespans.

Study three further uses the commercial contact networks to investigate the role of multiple transmission pathways on the genetic relatedness of 167 *C. jejuni* isolates sampled from across 30 commercial poultry farms. Permutational multivariate analysis of variance and distance-based linear models were used to explore the relative importance of network distances as potential determinants of the pairwise genetic relatedness between the *C. jejuni* isolates, with study results highlighting the importance of transporting feed vehicles in addition to the geographical proximity of farms and the parent company in the spread of disease.

In the last of the four *C. jejuni* studies, a compartmental disease transmission model was developed to simulate both the spread and sequence mutations across an outbreak within the commercial poultry industry. Simulated sequences were used in an analysis mirroring the methods used in study three in order to validate the approaches examining the contribution of local contacts and networks contacts towards disease transmission. An additional analysis is also performed in which the simulated sequence data is used to infer a transmission tree and explore the use of pathogen phylogenies in determining who-infected-whom across different model systems.

A further study, motivated by the application of whole genome sequence data to infer transmission, investigated the spread of *S. aureus* within the New Zealand dairy industry. This study demonstrated how whole-genome sequence data can be used to investigate pathogen population and evolutionary dynamics at multiple scales: from local to national and international. For this study, the genetic relatedness between 57 bovine derived *S.*

aureus isolates sampled from across 17 New Zealand dairy herds were compared with 59 *S. aureus* isolates that had been previously sampled and characterised from humans and domestic pets from across New Zealand and 103 *S. aureus* isolates extracted from GenBank that included both human and livestock isolates sampled from across 19 countries. Results from this study not only support evidence showing that the movement of live animals is an important risk factor for the spread of *S. aureus*, but also show that using cattle-tracing data alone may not be enough to fully capture the between farm transmission dynamics of *S. aureus*.

Overall, by using these two pathogen examples, this thesis demonstrates the potential use of pathogen whole-genome sequence data alongside contact network data in an epidemiological investigation, whilst highlighting the limitations and future challenges that must be considered in order to continue to develop robust methods that can be used to reliably infer the transmission and evolutionary dynamics across a range of infectious diseases.

Acknowledgements

My time at the EpiCentre and in Palmerston North has given me so much more than just a PhD. During my studies, I have had the privilege of being surrounded by some truly amazing people who continue to challenge and inspire me, and to whom I will be forever grateful. First, I would like to thank my supervisors; Carolyn Gates, Nigel French and Ji Zhang, who took a chance on me all those years ago and have continued to provide support, encouragement and guidance ever since. To all my supervisors, words cannot express my gratitude for all your contributions, the project would not be as it stands today if it wasn't for your knowledge and input.

I would also like to extend a massive thank you to all the organisations that have provided this project with funding including the New Zealand Biological Heritage National Science Challenge, the Poultry Industry Association of New Zealand, the Egg Producers' Federation, and Massey University's School of Veterinary Science as well as those organisations who allowed me access to data including TradeMe®, DairyNZ, and the Livestock Improvement Corporation. This project would not only not have been possible without you, but your support has also provided me with the opportunity to attend international conferences and workshops that have enriched my time working on the project.

I am also grateful to the many collaborators who have given their time and experience to the project. To Marti Anderson, thank you for sharing your statistical knowledge and advice, I hope someday to be half the statistically genius that you are. To Kerry McQueen and Michael Brooks from the New Zealand Poultry Industry Association and Thomas

Rawdon at Ministry for Primary Industries, thank you all for your patience with me as I tried to navigate the New Zealand poultry industry; your expertise and feedback has been invaluable to the project. The same goes for Scott McDougall and his team at Cognosco, thank you all for sharing your knowledge on the dairy industry and bovine mastitis in New Zealand; no matter my question, you were always able to provide an answer. I also massively appreciate and respect all the New Zealand farmers, both in the poultry industry and dairy industry, who participated in the project and gave us permission to collect samples; without your research such as this would not be possible and I am truly indebted to you all.

I am also hugely grateful to all the staff in both the Epicentre and the "EpiLab who I have had the privilege to work alongside. In particular I would like to thank Anne Midwinter, David Wilkinson, and Ahmed Fayaz, in addition to Deborah Williamson at the University of Melbourne, who completed all the sampling and lab work for the project; I would not have had any data to work with if it wasn't for you all. To everyone at the Epicentre; Cord, Naomi, Emilie, Chris, and Art, I want to thank you all for not only your help and encouragement, but also for making the office such a pleasure to have worked in by taking the time to always ask about my day and having a chat in the tearoom. I also massively appreciate all the work and support provided by Simon and Christine; I really do not know if the EpiCentre could function without you both. Thank you, Simon, for also always reminding us to take a break and switch-off, keeping us fuelled with morning coffee, and fixing everyone's computers (including mine on numerous occasions). You have always been willing to lend an ear to my worries and complaints, and I really appreciate all that you have done.

I would also like to say a big thank you to all the students, past and present, that have shared in all the highs and lows of completing a PhD. Many of you, and your families, have welcomed me into your life with nothing but smiles on your faces and I have truly

appreciated the time you have given me. I will always remember my first Kiwi potluck after being in New Zealand for only a few days, although hosted by two French people; Nelly and Matthieu. I want to thank you both for your friendship and the many good nights, and early mornings, spent in good company chatting over a great plate of food. Sebou, you were also a great support when I first arrived and paved the way for many adventures in the New Zealand wilderness.

Katja, thank you for always being there to share a good bottle of wine with me and have a good natter. I am so grateful to have had you just down the hall from my office; our lunch time chats together have always been a welcomed relief and I hope we can go out for many more lunches in the future. Caitlin, we also shared many great bottles of wine together and I will never forget the cheese and wine night you so fantastically hosted. Thank you for all your words of advice and help with post-doc applications, and for introducing me to riding again; I will always treasure our trip up to Rotorua and maybe someday we will go on a hack together again. Alicia, your energy, craziness and laughter always made sure we had a good time together and I will always be grateful for the advice you gave me before arriving in New Zealand; it was because of you I found the house up on Summerhill. Masako, thank you for your enthusiasm and kindness; you have an amazing sense to grasp at every opportunity and make the most of every moment often bringing me and many other people along on the ride. Milan, thank you for all the times we spent pouring over bioinformatics and making me feel as if I was not the only one struggling with many of its complexities. You and Prakriti have a beautiful family and I always enjoyed seeing you all and sharing in more good food together. Yuni, thank you for bringing your positivity into the office; I never saw you without a smile on your face and it was great to be able to share in your culture. Hopefully one day I will be able to visit you in Indonesia. Meow, you are also someone who is always thinking of other and has a smile to give. I will always remember your birthday crossword and our time on the mountain together as newbies. Diana, thank you for all our brunches together; I have always felt better after our talks

over coffee and cheesecake. Lukas, we also indulged in a lot of coffee and cake together and I will also remember our walks in the Esplanade; thank you for always being there. Omar, I also have many treasured memories of our time together, although Vinegar Hill still remains hazy. You introduced me to good tequila and Latin dancing, a great mix, and I hope one day we'll both make it to Canada and catch -up like old times. Jason, your time here was only brief, but I am so grateful that you decided to come to New Zealand, and I hope I have a friend for life. I will always remember our time in Chang Mai, drinking cocktails with Lena and dancing into the night. Marie, thank you for always being there to offer a good hug and some words of wisdom; it was a shame that our offices were so far apart. Reed, thank you for always sharing in my frustration over coding and being the best drinking partner on crate day. Lily, thank you for organising the writing boot camps for which I owe many words of this thesis to. How you juggle all your work with family and still have time to offer others your support is truly incredible, and I have enjoyed every moment we've spent together either in the gym or at Massey. I feel extremely lucky to have met each and every one of you for you have all contributed to the many wonderful experiences I have had during my studies and without your support, it would have been a lot harder to settle into a new life in New Zealand after moving away from family and starting my PhD.

I would like to extend a special thanks to some truly incredible people whose friendship and support has helped shaped me into who I am today. First, Guilhem and Lei, out of everyone I know, you two have the biggest hearts and so much love to give, and throughout my last year your kindness and support has meant the world to me. Guilhem, I will be forever grateful to have met you and cherish our friendship greatly. I remember many late nights talking over a glass of whiskey and sharing our pains and our fortunes. Lei, I am so happy that you and Guilhem found each other and since we've meant you have always taken care of me; making sure I'm doing alright and keeping very well fed. I

am so grateful for all the time you have made me welcome in your home and I will never forget the time you taught me how to make dumplings.

To Jun and Arata, I could not think of any better people to share an office with, and I only hope that someday I will be as good as an epidemiologist as you both. Throughout my project, you have always had time to answer my question, discuss my work, and help with my code; making me a better researcher, but much more than that, you both have helped me to overcome many of my physical and mental barriers. Arata, you are a superhuman and have an incredible gift in lifting people up and helping them push past their own boundaries, including myself, for which I will be forever grateful. Jun, you are an amazingly giving person who puts their heart and soul into everything they do. You have always been there when I have felt like giving up and have known exactly what to say to get me back up and ready to face the next challenge. I will never forget all the hours we all spent doing GRIT and ZUU together. My time here would truly not have been the same without you both. I also was lucky enough to find another family in the gym; to Janine, Mark, Masood, and Liam, all of whom have helped me train, I would like to thank you all for also inspiring me to keep on pushing myself and although there have often been times where I have cursed you, I have thoroughly enjoyed every minute spent training with you all. I would like to say a massive thanks to Gabriel, I will always remember the day we met, and I still think there is no better way to finish a long tramp than with a fresh apple crumble and ice cream. I treasure all the tramps we had since then, even the one up to sunrise hut, and want to thank you for reminding me not only what I should expect from other people but also what I should expect from myself. I will never forget our time diving in Fiji, and I hope someday we can be dive buddies again. To Gemma, you have always shown an amazing passion for your work both throughout the time we did our MSc together and during our PhDs. Both you and Rob were there for me at a time when I was close to quitting, and I would like to thank you both for your encouragement and support. I am very grateful you made it to New Zealand, although your time here was far too brief;

your passion for work is infectious and you have shown me just what is cable of ourselves if we set our minds to it. To Hamish and Sarah, my flatmates over the past three years, you have been my home away from home with Rascal and Ziggy and it has been lovely knowing that every day I can leave my work in the office and return to a house full of friendly faces. I hope we make many more memories up North.

There are also a few people, without whom, I would definitely not be where I am today. I am extremely lucky to have a loving family who have continued to support me even from the other side of the world. To Charlotte and Emma, you may not be related but you are still family. Thank you for coming all the way to visit me and spending Christmas tramping up mountainside and erecting tents in the rain; it was some of the best time I've had in New Zealand. To my Mum and Dad, you are the ones who built the foundations that make me who I am, and I thank you for always encouraging me to pursue a career in science. I have appreciated every phone call and Skype conversation that we've had whilst I've been in New Zealand, even when you've forgotten about the time difference and its been 3am in the morning, as I know you will always be there for me. To my Grandma and Pop, you have also been a massive presence in my life and have never failed to be there for me. I will always cherish the moments we spent together whilst travelling on your visit to New Zealand. To my brother, Dean, without whom I wouldn't even have had the funds to get myself to New Zealand, you are not a man of many words but I know you would do anything for me; thank you for always being there when I need you. To my twin sister, Davina, we have always been more than just sisters and you continue to take care of me more than anyone else despite being worlds apart. I've appreciated every food parcel and every WhatsApp message; I can't imagine my life without.

To end I would like to dedicate this thesis to Alisdair. You have stuck by me through all my setbacks, have listened to all my tears and tantrums, and have been there on all the days when I have just wanted to give up. In the times when I've struggled to believe in

myself, you have given me the courage and strength to keep on chasing my dreams. You continue to inspire me to be the best I can be, and I am excited to start the next chapter of our lives together; with many more adventures to come !

Publications and Presentations

Publications

Greening SS., Mulqueen K., Rawdon TG., French NP., Gates MC. Estimating the level of disease risk and biosecurity on commercial poultry farms in New Zealand. *N. Z. Vet. J.* (In Press) 2020

French NP., Zhang J., Carter GP., Midwinter AC., Biggs PJ., Dyet K., Gilpin BJ., Ingle DJ., Mulqueen K., Rogers LE., Wilkinson DA., Greening SS., Mullner P., Williamson DA. Rapid and concurrent emergence of fluoroquinolone and tetracycline-resistant *Campylobacter jejuni* in humans and poultry in New Zealand. *Emerg. Infect. Dis.* 25(12):2226-2234, 2019

Pannell JL., Dencer-Brown AM., Greening SS., Hume EA., Jarvis RM., Mathieu C., Mugford J., Runghen R. An early career perspective on encouraging collaborative and interdisciplinary research in ecology. *Ecosphere*. 10(10): e02899, 2019

Gates C., Greening SS., Han J-H., Hidano A. Tracking the spread of infectious livestock diseases through their molecular footprints. *Vetscript* 30(5); 10-11, 2017

Presentations

Greening SS., Zhang J., French NP., Gates C. *Pathogen genomics: a new steppingstone in epidemiological investigations*. One Health Aotearoa Symposium. Wellington, New Zealand, 2019

Greening SS., Zhang J., French NP., Gates C. *Untangling the complex relationships between host population contact structure, disease transmission dynamics, and pathogen evolution*. The 15th International Symposium of Veterinary Epidemiology and Economics. Chiang Mai, Thailand, 2018

Greening SS., Wilkinson D., Midwinter AC., Williamson D., Zhang J., Anderson MJ., Gates C., French NP. *Transmission dynamics of Campylobacter jejuni (ST-6964) in New Zealand's commercial poultry network.* Australian and New Zealand College of Veterinary Scientists Science Week. Gold Coast, Australia, 2018

Greening SS., Zhang J., French NP., Gates C. *Social network analysis in the genomic era.* NetSci. Paris, France, 2018

Greening SS., Zhang J., French NP., Gates C. *Bridging the gap: linking molecular and population contact network data.* New Zealand Veterinary Association Conference; All Creatures Great and Small. Hamilton, New Zealand, 2018

Greening SS., Mulqueen K., Rawdon TG., French NP., Gates C. *Infectious disease risk pathways and their influence on biosecurity practices in the New Zealand commercial poultry industry.* School of Veterinary Science Post Graduate Colloquium. Palmerston North, New Zealand, 2018

Greening SS., Zhang J., French NP., Gates C. *Biosecurity Network Interventions in New Zealand's Livestock networks.* BioHeritage Early Career Workshop. Lincoln, New Zealand, 2018

Greening SS., Zhang J., French NP., Gates C. *Pathogens in poultry; the hidden influence of contact networks.* The annual meeting of the New Zealand Food Safety Science and Research Centre. Nelson, New Zealand, 2017

Greening SS., Wilkinson D., Midwinter AC., Williamson D., Fayaz A., Zhang J., Gates C., French NP. *Integrating molecular data with livestock contact networks to support inferences on pathogen transmission and evolutionary dynamics.* One Health Aotearoa Symposium. Wellington, New Zealand, 2017

Contents

Abstract	i
Acknowledgements	v
Publications and Presentations	xiii
Contents	xv
Abbreviations	xxi
List of figures	xxvii
List of tables	xxxix
Preface	xliv
 Chapter 1: Introduction	 1
1.1. Infectious disease dynamics in a connected world	2
1.2. Host contact networks and pathogen phylogenetics	3
1.3. Research focus	4
1.3.1. New Zealand's commercial poultry industry	6
1.3.2. Epidemiology of <i>C. jejuni</i> in New Zealand	7
1.3.3. New Zealand's dairy cattle industry	10
1.3.4. Epidemiology of <i>S. aureus</i> in New Zealand	13
1.4. Thesis structure	16
1.5. References	18

Chapter 2: Literature Review	25
2.1. Abstract	26
2.2. Introduction	28
2.3. Population structure, networks, and disease dynamics	31
2.3.1. Network theory	32
2.3.2. SNA and infectious disease dynamics	35
2.4. Pathogen population dynamics and infectious diseases	38
2.4.1. Pathogen phylogenetics	44
2.4.2. Genealogy-based modelling	46
2.5. Epidemiological models incorporating genetic data	57
2.5.1. Evolution of contact networks and pathogen traits	51
2.6. Current limitations and future challenges	52
2.6.1. Missing or imperfect contact network data	52
2.6.2. Modelling dynamic networks	53
2.6.3. Phylogenetic complexities	54
2.7. Conclusion	57
2.8. Acknowledgements	57
2.9. References	58

Chapter 3: Estimating the level of disease risk and biosecurity on commercial poultry farms in New Zealand

Chapter 3: Estimating the level of disease risk and biosecurity on commercial poultry farms in New Zealand	81
3.1. Abstract	82
3.2. Introduction	84
3.3. Materials and Methods	86
3.3.1. Survey implementation	86
3.3.2. Survey design	87
3.3.3. Data processing	88
3.3.4. Generating the farm-level contact risk score	89
3.3.5. Generating the farm-level biosecurity score	92
3.3.6. Statistical analysis	93
3.4. Results	94
3.4.1. Poultry demographics	94
3.4.2. Contact patterns and contact risk scores	94
3.4.3. Biosecurity practices and frequency scores	100
3.4.4. Association between risk and biosecurity scores	102
3.4.5. Producer concerns about poultry health	102

3.5. Discussion	107
3.6. Conclusion	113
3.7. Acknowledgements	114
3.8. References	115

Chapter 4: Characterising the disease risk from backyard poultry trade networks and migratory birds to New Zealand’s commercial poultry industry..... 121

4.1. Abstract	122
4.2. Introduction	124
4.3. Materials and Methods	127
4.3.1. Commercial poultry network	127
4.3.2. Backyard poultry network	129
4.3.2.1. TradeMe® transactions	129
4.3.2.2. Backyard poultry networks	130
4.3.2.3. Backyard poultry simulation model	132
4.3.3. Migrating wilds birds and waterfowl	133
4.3.4. Transmission between commercial and non-commercial poultry	135
4.4. Results	136
4.4.1. Commercial poultry network	136
4.4.2. Backyard poultry network	139
4.4.2.1. TradeMe® transactions	139
4.4.2.2. Backyard poultry networks	139
4.4.2.3. Backyard poultry simulation model	146
4.4.3. Migrating wilds birds and waterfowl	146
4.4.4. Transmission between commercial and non-commercial poultry	150
4.5. Discussion	153
4.6. Conclusion	161
4.7. Acknowledgements	162
4.8. References	163

Chapter 5: Transmission dynamics of *Campylobacter jejuni* (ST-6964) in New Zealand's

commercial poultry network	169
5.1. Abstract	170
5.2. Introduction	172
5.3. Materials and Methods	175
5.3.1. Sample collection, isolate culture and WGS	175
5.3.2. Genome assembly and wgMLST analysis	176
5.3.3. Comparative genomics of mobile elements	177
5.3.4. Genetic distance matrices	179
5.3.5. Model matrices	179
5.3.6. Relating genetic distances and model matrices	183
5.3.7. Explaining variation in genetic distances	184
5.4. Results	185
5.4.1. Sample collection, isolate culture and WGS	185
5.4.2. Whole-genome MLST analysis	186
5.4.3. Comparative genomics of mobile elements	190
5.4.4. Characteristics of model matrices	190
5.4.5. Relating genetic distances and model matrices	194
5.4.6. Explaining variation in genetic distances	196
5.5. Discussion	198
5.6. Conclusion	201
5.7. Acknowledgements	202
5.8. References	203

Chapter 6: Assessing the impact of host contact network structure on the phylogeny of *Campylobacter jejuni* in commercial broiler systems: a simulation modelling approach

	211
6.1. Abstract	212
6.2. Introduction	214
6.3. Material and Methods	217
6.3.1. Disease simulation model	217
6.3.2. Model demographics	217
6.3.3. On-farm disease dynamics	221
6.3.4. Between-farm disease dynamics	224
6.3.4.1. Spread through network contacts	224
6.3.4.2. Spread through local contacts	226

6.3.5. Estimating the probability of infection	227
6.3.6. Sequence simulation	229
6.3.7. Simulations conditions	230
6.3.8. Phylogenetic analysis	231
6.3.9. Statistical analysis	232
6.4. Results	234
6.4.1. Industry demographics and contacts	234
6.4.2. Disease transmission model	238
6.4.3. Phylogenetic and statistical analyses	240
6.5. Discussion	243
6.6. Conclusion	249
6.7. Acknowledgements	249
6.8. References	250

Chapter 7: Genetic relatedness of *Staphylococcus aureus* isolates obtained from New Zealand dairy cattle

7.1. Abstract	260
7.2. Introduction	262
7.3. Materials and Methods	266
7.3.1. Current study: sampling, microbiology and WGS	266
7.3.2. Genomic analyses: current study	267
7.3.3. Antimicrobial sensitivity	268
7.3.4. <i>S. aureus</i> within New Zealand	268
7.3.5. <i>S. aureus</i> ST-1 globally	269
7.3.6. Cattle trade network	270
7.3.7. Antimicrobial usage	271
7.3.8. Permutational multivariate analysis of variance	274
7.4. Results	274
7.4.1 Genomic analysis: current study	274
7.4.2. Antimicrobial sensitivity	282
7.4.3. <i>S. aureus</i> within New Zealand	286
7.4.4. <i>S. aureus</i> ST-1 globally	289
7.4.5. Cattle trade network	292
7.4.6. Antimicrobial usage	293
7.4.7. Permutational multivariate analysis of variance	295

7.5. Discussion	296
7.6. Conclusion	301
7.7. Acknowledgements	302
7.8. References	303
 Chapter 8: General discussion	313
8.1. Introduction	314
8.2. Overview of results: <i>Campylobacter jejuni</i>	315
8.3. Overview of results: <i>Staphylococcus aureus</i>	318
8.4. Future opportunities	321
8.5. Data limitation	324
8.6. Developing methodologies.....	326
8.7. Concluding comments	327
8.8. References	328
 Appendix A: Supplementary Material - Poultry Industry Survey Questionnaire 2016	335
 Appendix B: Supplementary Material - Chapter 3	349
 Appendix C: Supplementary Material - Chapter 4	367
 Appendix D: Supplementary Material - Chapter 5	381
 Appendix E: Supplementary Material - Chapter 7	403
 Appendix F: DRC16 Statement of Contribution forms	421

Abbreviations

AI	Avian Influenza
AIC	Akaike Information Criterion
AMR	Antimicrobial Resistance
AMU	Antimicrobial Usage
ANOVA	Analysis of Variance
BAG	Biofilm-Associated Genes
bp	Base Pairs
°C	Degrees Celsius
Cef	Cefoxitin
<i>C. jejuni</i>	<i>Campylobacter jejuni</i>
CARD	Comprehensive Antibiotic Resistance Database
CC	Clonal Complex
CI	Confidence Interval
CJEI	<i>Campylobacter jejuni</i> integrated elements
CLSI	Clinical and Laboratory Standards Institute
dbRDA	Distance-Based Redundancy Analysis
DD	Daily Dose
<i>df</i>	Degree of Freedom
DistLM	Distance-Based Linear Models
DNA	Deoxyribonucleic Acid

E. coli *Escherichia coli*
 ef Equal Base Frequencies
 EHF Ebola Haemorrhagic Fever
 EPF Egg Producers' Federation
 ESR Institute of Environmental Science and Research
 F81 Felsenstein Model
 F84 Hasegawa-Kishino-Yano Model
 FMD Foot-and-Mouth Disease
 GC Guanine-Cytosine
 GDP Gross Domestic Product
 GeP Genome Profiler
 GHE Global Health Estimates
 GO Gene Ontology
 GPS Global Positioning Systems
 GSCC Giant Strongly Connected Component
 GTR General Time Reversible Model
 GWCC Giant Weakly Connected Component
 HIV Human Immunodeficiency Virus
 HPAI Highly Pathogenic Avian Influenza
 I Intermediate
 iTOL Interactive Tree of Life
 JC Jukes and Cantor
 K80/1 Kimura Model
 KDE Kernel Density Estimate
 LA-MRSA Livestock-Associated Methicillin-Resistant *S. aureus*
 LIC Livestock Improvement Corporation

<i>M. bovis</i>	<i>Mycoplasma bovis</i>
mCCDA	Modified Charcoal Cefoperazone Deoxycholate Agar
^m EpiLab	Molecular Epidemiology and Public Health Laboratory
mg	Milligrams
MIC	Minimum Inhibitory Concentrations
ML	Maximum Likelihood
ml	Millilitre
mm	Millimetre
MLST	Multilocus Sequence Typing
<i>m</i> MDS	Metric Multidimensional Scaling
MPI	Ministry of Primary Industries
MRCA	Most Recent Common Ancestor
MRSA	Methicillin-Resistant <i>Staphylococcus aureus</i>
MSEI	Maternally-Susceptible-Latent-Infectious
MSS	Mean Sum-of-Squares
MSSA	Methicillin-Susceptible <i>Staphylococcus aureus</i>
MST	Minimum Spanning Tree
NA	Not Applicable
NAIT	National Animal Identification and Tracing
NCBI	National Centre for Biotechnology Information
NiV	Nipah Virus
Nov	Novabiocin
NZBH	New Zealand Biological Heritage
NZD	New Zealand Dollar
NZFSA	New Zealand Food Safety Authority
OIE	World Organisation for Animal Health

OXAOxacillin

PCParent Company

PCCPearson’s Correlation Coefficient

PenPenicillin

PERMANOVAPermutational Analysis of Variance

PFGEPulsed-Field Gel Electrophoresis

PIANZPoultry Industry Association of New Zealand

RResistant

R_0 Basic Reproduction Number

RFIDRadio Frequency Identification Device

ρ Spearman’s Rank Correlation

rRNARibosomal Ribonucleic Acid

SSensitive

S. aureus*Staphylococcus aureus*

S. dysgalactiae*Streptococcus dysgalactiae*

S. uberis*Streptococcus uberis*

SAMMSeasonal Approach to Managing Mastitis

SARSSevere Acute Respiratory Syndrome

SCCStrongly Connected Components

SEStaphylococcal Enterotoxins

SISusceptible-Infectious

SIRSusceptible-Infectious-Recovered

SNASocial Network Analysis

SNPSingle Nucleotide Polymorphism

SPShortest Path

SSSum-of-Squares

SSTI Skin and Soft Tissue Infections

ST Sequence Type

STEC Shiga Toxin-Producing *Escherichia coli*

STI Sexually Transmitted Infections

SYM Symmetrical Model

TA Territorial Authorities

Tet Tetracycline

TIM Transitional Model

TMRCA Time to the Most Recent Common Ancestor

TrN Tamura-Nei Model

TVM Transversion Model

uf Unequal Base Frequencies

µg Micrograms

UK United Kingdom

UPGMA Unweighted Pair-Group Method with Arithmetic Mean

US United States

USDA United States Department of Agriculture

VTEC Verotoxigenic *Escherichia coli*

WCC Weakly Connected Components

wgMLST Whole Genome Multilocus Sequence Typing

WGS Whole Genome Sequencing

WHO World Health Organisation

WWOOF Willing Workers on Organic Farms

XLN Ceftiofur

List of Figures

- 1.1.** Schematic diagram showing the structure of the New Zealand poultry industry including breeder farms and hatcheries (green), broiler farms (purple), and layer farms (blue) **p.8**
- 1.2.** A typical annual production cycle for New Zealand dairy farms using a pasture-based system with seasonal calving taking place in the spring (purple; July-October), followed by mating (yellow; October-December), lactation and drying off (green; October-June), and weaning and pregnancy testing (blue; February-April). **p.11**
- 2.1.** Directed network graph $G = \{V, E\}$ showing a set of vertices otherwise known as nodes or actors; A, B, C and, D connected by multiple edges each representing a different relationship including a unidirectional edge (ud), multiple edges (m), bidirectional edges (bd) and loops (L) **p.33**
- 3.1.** (a) The distribution of the 414 active poultry producers register with either the Poultry Industry Association of New Zealand (PIANZ) or the Egg Producers' Federation (EPF) as of June 2016, and (b) the subset of poultry producers who responded to the 2016 poultry survey including 24 breeder enterprises (light blue), 57 broiler enterprises (pink), 33 layer enterprises (yellow), and 6 other poultry enterprise (navy) **p.95**
- 3.2.** Boxplot showing the estimated risk score for 120 poultry producers in New Zealand's commercial poultry industry including 33 layer enterprises, 57 broiler enterprises, 24 breeder enterprises and 6 enterprises representing either duck, turkey or pullet operations. The values in the graph represent the minimum, maximum, median, first quartile and third quartile in the data with the median value (\tilde{x}) printed below each plot **p.98**

- 3.3.** The reported frequency of implementing seven common biosecurity measures amongst 120 poultry producers in New Zealand's commercial poultry industry **p.101**

- 3.4.** Boxplot showing the estimated biosecurity score for 120 poultry producers in New Zealand's commercial poultry industry including 33 layer enterprises, 57 broiler enterprises, 24 breeder enterprises and 6 enterprises representing either duck, turkey or pullet operations. The values in the graph represent the minimum, maximum, median, first quartile and third quartile in the data with the median value (\bar{x}) printed below each plot **p.103**

- 3.5.** Scatter plot showing the relationship between the estimated biosecurity score, calculated from the reported frequency of implementing seven common biosecurity practices, and the estimated disease risk score for 120 poultry producers in New Zealand's commercial poultry industry including 33 layer enterprises, 57 broiler enterprises, 24 breeder enterprises and 6 enterprises representing either duck, turkey or pullet operations. Point circumference is proportional to farm frequency whilst the pie chart shows the breakdown of farms by production type. Pearson's correlation coefficient (PCC) has been calculated and 95% confidence intervals are shown for a p -value <0.01 **p.104**

- 4.1.** Schematic diagram showing the construction of the commercial poultry network with (a) showing three farms connected to transporting vehicles belonging to different companies: one providing feed, the other removing waste. All three farms use the waste company (blue lines) whilst only Farms 2 and 3 use the feed company (orange lines) with those farms linked to the same company being directly linked in (b) such that farms 1 is connected to farms 2 and 3 via the movement of waste, and farms 2 and 3 are connected via the both the movement of feed and waste **p.129**

- 4.2.** (a) Location of commercial poultry enterprises in New Zealand registered with either PIANZ or EPF as of June 2016, (b) choropleth map showing the density of commercial poultry enterprises within the 67 territorial authorities and (c) commercial contact network for the on – and off-farm **p.138**

movement of feed, live birds and hatching eggs and, poultry waste and litter

- 4.3.** Frequency of (a) chicken, (b) bantam, (c) duck and (d) all poultry trades occurring through the online auction website TradeMe® in New Zealand by week of year from 01st January 2012 to 31st December 2018 **p.141**

- 4.4.** Trader contact networks for (a) chicken*, (b) bantam, (c) duck and (d) all poultry* trades occurring through the online auction website TradeMe® in New Zealand from 01st January 2018 to 31st December 2018. Node colour and size highlights those nodes with a degree centrality measure ≥ 50 for bantam and duck trades (red: degree ≥ 50 , orange: degree < 50) and ≥ 100 for chicken and all poultry trades (red: degree ≥ 100 , orange: degree < 100). To improve clarity in the visualization, only connections to and from nodes with a degree ≥ 100 are shown in the networks showing (a) chicken and (d) all poultry trades **p.145**

- 4.5.** Spatial networks for (a) chicken, (b) bantam, (c) duck and (d) all poultry trades occurring through the online auction website TradeMe® in New Zealand from 01st January 2018 to 31st December 2018. Edge colour indicates the frequency of trades (*i.e.*, the number of days between two consecutive trades going to and from the same two nodes) **p.147**

- 4.6.** Choropleth maps showing the number of bird sighting reported to eBird in each of the 67 territorial authorities from 01st January 2018 to 31st December 2018 with (a) indicating the total number of bird species reported, (b) the number of resident wild birds that are known to migrate within New Zealand, (c) the number of exotic wild birds known to migrate to and from New Zealand, and (d) the number of resident water birds within the Anatidae family. The raw point data is presented alongside each map, with the choropleth maps adjusted for by the population density in each region **p.149**

- 4.7.** The spatial overlap between (a) the commercial poultry network (yellow edges) and the backyard poultry trade network (red edges) with the locations of commercial poultry producers indicated by the blue node, and bivariate **p.151**

choropleth maps showing further overlap between (b) commercial poultry enterprise (x-axis) and backyard poultry trade (y-axis), (c) resident migratory wild birds (x-axis) and backyard poultry trade (y-axis), (d) exotic migratory wild birds (x-axis) and backyard poultry trade (y-axis), (e) commercial poultry enterprises (x-axis) and resident migratory wild birds (y-axis), and (f) commercial poultry enterprises (x-axis) and exotic migratory wild birds (y-axis). All data extracted from 01st January 2018 to 31st December 2018

- 4.8.** Bivariate choropleth maps showing the spatial overlap between resident water birds in the Anatidae family (y-axis) and (a) commercial poultry enterprises (x-axis), (b) backyard poultry trade (x-axis), (c) exotic migratory wild birds (x-axis), and (d) resident migratory wild birds. All data extracted from 01st January 2018 to 31st December 2018 **p.152**

- 5.1.** Schematic diagram showing the construction of (a) bimodal and (b) unimodal network graphs. In network (a) there are two node types representing either farms or transporting companies moving on- and off-farm. In network (b) the farms have been directly linked if they have a transporting company in common. The shortest path (SP) between any two farms is the least number of steps between the farms. For example, network (b) shows two paths between farm 2 and farm 3. The dashed line connects the two farms via farm 1 whilst the bold line is a direct link between the two farms, making it the shortest path. (c) The shortest path matrix for transport company C (green lines) with 1 indicating there is a direct link between two farms in (b) and 99 indicating no link **p.181**

- 5.2.** (a) Map showing the geographical distribution of New Zealand commercial poultry premises (n = 30) with sampled isolates positive for *C. jejuni* ST-6964 (n = 167) alongside (b) a phylogeny showing the genetic relatedness of the isolates based on core polymorphic loci. Production type is indicated by the shape of the terminal node (triangles: poultry breeder and circles: poultry grower) with the three colours indicating the three different parent companies to which the farm belongs. The figure has been created using the online tool Microreact (Argimón *et al.* 2016) **p.187**

- Two-dimensional metric multidimensional scaling (*m*MDS) ordination plots based on the allelic dissimilarity matrix between 167 *C. jejuni* isolates with isolates coloured according to (a) the parent company and (b) the network community to which they belong **p.188**
- 5.3.**
- A tanglegram of the maximum-likelihood tree showing the population structure of 167 *C. jejuni* ST-6964 isolates rooted by the outgroup (left) compared to a dendrogram representing the Euclidean distances between the sites from which isolates were sampled (right) with line colour indicating one of three poultry suppliers (A, B or C) with farms belonging to supplier A located in two geographical regions (regions 1 and 2) in comparison to poultry suppliers B and C whose farms are geographically clustered in one region. To maximise tree congruence a two-tree crossing minimization technique was used based on a greedy forward selection algorithm **p.189**
- 5.4.**
- A rooted maximum-likelihood phylogenetic tree reconstructed from core polymorphic loci between 167 *C. jejuni* ST-6964 sampled from 30 New Zealand commercial poultry farms. The coloured rings indicate (i) the parent company (PC) of the farm from which the isolate was collected, (ii) the farm production type (PT) (broiler versus breeder), (iii) the presence or absence of *C. jejuni* integrated element 1 (*CJIE1*), (iv) *CJIE1v*, (v) *CJIE2*, (vi) *CJIE3*, and (vii) *CJIE4*. The figure has been created using the online tool Interactive Tree of Life (iTOL) (v4.4.2) **p.191**
- 5.5.**
- A rooted maximum-likelihood phylogenetic tree reconstructed from core polymorphic loci between 167 *C. jejuni* ST-6964 sampled from New Zealand commercial poultry farms (n = 30). The coloured rings indicate (i) the parent company (PC) of the farm from which the isolate was collected (ii) the farm production type (PT) (broiler versus breeder), (iii) the presence or absence of a *Tet*-like plasmid that has been associated with resistance (*pTetO*), (iv) the presence or absence the virulent plasmid *pVir*, and the selected truncation of two functional genes (v) *amtB* encoding an ammonium transporter and (vi) *glcD* encoding **p.192**
- 5.6.**

the putative glycolate oxidase subunit D. The figure has been created using the online tool Interactive Tree of Life (iTOL) (v4.4.2)

Second-stage metric multidimensional scaling (*m*MDS) ordination plots showing proximities (rank matrix correlations) between the allelic dissimilarity matrix (“allelic”) and each of several model distance matrices: “euc” = Euclidean distance; “road” = road distance; “feed” = feet network distance; “live” = live birds network distance; “waste” = waste and litter network distance; “parent” = parent company distance. Superimposed on each plot is the minimum spanning tree (MST). The *m*MDS plot shown in (a) includes parent company whereas (b) does not include parent company

5.7. **p.195**

Schematic diagram showing an overview of the *Campylobacter* transmission model among three individual farms at time t_n . Each farm has a fixed number of sheds housing a fixed number of birds with the shed colour indicating its infectious status (yellow: infectious and blue: non-infectious). When *Campylobacter* first enters a shed, the within shed prevalence of *Campylobacter* for t_{n+1} is modelled using an MSEI compartment model (1) with the number of infected birds recorded in each time step. The within shed prevalence is used to determine the likelihood of transmission between sheds in the same flock (2) indicated by the vertical arrows in farm 1 where transmission has occurred from shed 1 to shed 2 and from shed 2 to sheds 3 and 4. Transmission between flocks (3) can occur via two mechanism (3a) indicates spread from farm 1 to farm 3 due to a direct contact in the network showing the movement of transporting vehicles (horizontal arrow) whilst (3b) indicates spread from farm 1 to farm 2 due the spatial proximity of the production premises (dashed boxes)

6.1. **p.218**

Schematic showing the MSEI deterministic compartmental model where solid horizontal lines indicate the movement of birds between compartments whilst dashed vertical lines represent birds leaving the compartment due a background mortality rate (α). In the model all 1-day-olds chicks start immune (M) with maternal antibodies decaying at a constant

6.2. **p.222**

rate (γ) ensuring that all chicks become susceptible (S) within 2-3 weeks. If infected one susceptible bird was moved to the exposed group (E) after which, each susceptible individual in the infected shed received the force of infection (λ) at a rate that varied as a function of the number of infectious individuals in the shed; such that, $\lambda = \lambda(I)$. Once exposed, birds transitioned into the infectious group (I) at a constant rate (δ) reflecting a latency period of 1.5 days in which exposed birds remained infected but not yet infectious and so therefore did not contribute to the force of infection. Birds in the infected group (I) were assumed to be persistently infected for the remaining duration of the production cycle

6.3. A schematic diagram showing a network graph (left) that could have been constructed from survey responses with vertices (V) belonging to different groups; for example, V1 (1, 2, 3 and 4) represent poultry farms whilst V2 (A, B, C and D) represent those transporting companies reported to have delivered goods or services to the connected farm. This graph can be used to construct the network on the right which contains just the vertices from the V1 group (*i.e.*, only the poultry farms that responded to the survey), with edges between vertices that share a common vertex belonging to the V2.group. For example, the green dashed edge connects farm 1 and farm 4 in the right-hand graph due to the movement of waste and litter shown in the left-hand graph by transporting company C p.225

6.4. A schematic diagram showing how to calculate the total probability (P) of a new infection occurring in an uninfected shed (*i.e.*, Shed 2, Farm 2: shown in blue) from an infected shed (shown in yellow). In this example no infection can result from a movement in either the waste network, or the live bird and hatching egg network as the farms do not share the same transporting companies. Therefore, both P_1 and P_3 are equal to 0. However, there is a probability of Shed 2 becoming from Farm 1, in which all shed are infected, due to either the close proximity of the two farms (indicated by the dashed lines) and the probability of a local contact (P_5), or via a contact in the feed network (P_2), as both Farms 1 and 2 share the same feed company. Shed 2 could also become infected via contact with the other infected shed on the farm (*i.e.*, Shed p.228

1, Farm 2). Therefore, the total probability of Shed 2 becoming infected is: $1 - [(1-P_1) \times (1-P_2) \times (1-P_3)]$

(a) A schematic diagram representing a transmission network containing three different transmission chains as indicated by the colour of the lines (blue, red and black). In this example, the primary case is Farm 1, who directly infects Farm 5 (blue chain), Farm 6 (red chain), and Farm 8 (black chain). There is then onward transmission from both Farms 5 and 6 as indicated by the direction of the arrows in each of the chains.

(b) shows the corresponding network that has been inferred using pathogen sequence data. In the network the direct

- 6.5.** transmissions between Farm 1 and Farm 6, Farm 1 and Farm 8, and Farm 7 and Farm 4 have been correctly identified, as indicated by the solid lines. Indirect transmission between Farm 1 and Farm 2 as well as between Farm 6 and Farm 7 has also been correctly identified, as indicated by the dashed line, meaning that these farms are not direct contacts in (a) but are in the same transmission chain. Therefore, the proportion of infectious contacts correctly identified is 42.9% (3/7), whilst proportion of contacts in a transmission chain correctly identified 71.4% (5/7) **p.233**

Histograms showing the pairwise Euclidean distance (km) between (a) all commercial broiler farms in New Zealand (n = 156 with 12,090 pairwise distances) with the arithmetic mean indicated by the dashed red line, and (b) the proportion of pairwise Euclidean distance within 7km of each other (480/12,090, 4.0%)

- 6.6.** **p.237**

Epidemic curves showing the proportion of all broiler farms (n = 156) that got infected on each day across all the simulations which the outbreak ran for the full 365 days including (a) 99 simulations with both local and network contacts, (b) 61 simulations with only network contacts, and (c) 96 simulations with only local contacts. The red line on each graph indicates the mean across all the model simulations

- 6.7.** **p.239**

Boxplot showing the cophenetic correlation coefficient over 99 model simulations including both network and local contacts (blue), 61 model simulations including only network

- 6.8.** **p.241**

contacts (yellow), and 96 model simulations including only local contacts (green) with values in the graph representing the minimum, maximum, median, first quartile and third quartile in the data

- 6.9.** Boxplot showing the Spearman's rank matrix correlation (ρ) between the model matrices for (a) the Euclidean distance between poultry farms, (b) the feed network, (c) the waste and litter network and, (d) the live bird and hatching eggs network and the genetic p -distance matrix between all simulated *C. jejuni* sequences sampled on the 365th day of each model simulation including 99 simulations with both network and local contacts (blue), 61 simulations with only network contacts (yellow), and 96 simulations with only local contacts (green). Values in the graph represent the minimum, maximum, median, first quartile and third quartile in the data **p.242**

- 7.1.** (a) shows one contrasting pair (blue branches; 1-0|0-1) whilst (b) shows another contrasting pair on the same tree (red branches; 1-1|0-0), however the maximum number of non-intersecting, contrasting pairs in the tree remains as one due to the common branches, highlighted in green, shared by both contrasting pairs. The “best” picking is the red pair with both the gene and the trait present whilst the “worst” picking is the blue pair. To handle confounding by lineage a pairwise comparisons algorithm can be used to identify the maximum number of non-intersecting, contrasting pairs in a tree and calculates the corresponding binomial test p -value taking into account the proportion of “best” and “worst” pickings over the set of non-intersecting, contrasting pairs **p.273**

- 7.2.** (a) Maximum-likelihood phylogeny generated from core single nucleotide polymorphisms across the 57 *S. aureus* isolates. The colour of the terminal tree nodes corresponds to the farm location shown in (b), and all isolates have been labelled with their sequence type identified by seven MLST genes. (b) Map showing the geographical distribution of selected dairy farms ($n=17$). The alphabetic labels (A-Q) have been used to identify the farms in the subsequent phylogenetic analysis. Figure created using the online tool Microreact (Argimón *et al.* 2016) **p.276**

- 7.3.** Resistance gene profiles of 57 *S. aureus* isolates presented alongside a maximum-likelihood tree generated from the core single nucleotide polymorphisms. Isolate IDs identify the date the sample was collected (dd/mm/yyyy), the farm from which it was collected from (A-Q) as indicated in Figure 7.2, and the animal ID number (####) **p.279**
- 7.4.** Virulence gene profiles of 57 *S. aureus* isolates presented alongside a maximum-likelihood tree generated from the core single nucleotide polymorphisms. Isolate IDs identify the date the sample was collected (dd/mm/yyyy), the farm from which it was collected from (A-Q) as indicated in Figure 7.2, and the animal ID number (####) **p.280**
- 7.5.** Antimicrobial sensitivity profiles of 50 *S. aureus* isolates presented alongside a maximum-likelihood tree generated from the core single nucleotide polymorphisms. Sensitivity was determined using a zone diffusion test following the procedures provided by the Clinical and Laboratory Standards Institute (CLSI) for penicillin (Pen), novobiocin (Nov), cefoxitin (Cef), tetracycline (Tet), ceftiofur (XNL), and oxacillin (OXA); with isolates being declared sensitive (S), intermediate (I) or resistant (R), based on CLSI recommendations. Isolate IDs identify the date the sample was collected (dd/mm/yyyy), the farm from which it was collected from (A-Q) as indicated in Figure 7.2, and the animal ID number (####) **p.283**
- 7.6.** Maximum-likelihood phylogeny generated from core single nucleotide polymorphisms across 57 *S. aureus* isolates from this study (*i.e.*, New Zealand dairy cattle) and 59 *S. aureus* isolates from a previous study by Grinberg and colleagues (Grinberg *et al.* 2010). Isolate IDs identify the year the sample was collected (yyyy) and the species the sample was taken from. The colour of the terminal tree node indicates the sequence type identified by seven MLST genes whilst the shape further identifies the species and site from which the isolate it was collected. BM: bovine milk, HC: human clinical case, HN: human nasal colonisation, CC: canine clinical case, CN: canine nasal colonisation, FC: feline clinical case and FP: feline perianal colonisation **p.287**

- (a) Maximum-likelihood phylogeny generated from core single nucleotide polymorphisms across 35 *S. aureus* ST-1 isolates from this study (*i.e.*, New Zealand dairy cattle) and the 103 *S. aureus* ST-1 isolates available in Genbank as of March 2019. Isolate IDs identify the year the sample was collected (yyyy) and the country from which it was sampled from whilst the shape of the terminal node indicates the species it was sampled from (circle = cattle, square = human, star = domesticated pets and, triangle = all other samples). (b) Map showing the countries from which the isolates were sampled from with points sitting on the centroid position. Point shape indicates the dominate species that has been sampled in the country
- 7.7.** **p.290**

List of Tables

2.1.	Network analysis glossary of terms for measures of centrality and cohesion	p.36
3.1.	The disease risk-ranking matrix showing the contribution of risk attributes within each of the eight-risk criterion to the estimated disease risk score	p.90
3.2.	Summary of demographic information amongst 120 producers in the New Zealand commercial poultry industry all of whom responded to the 2016 poultry survey	p.96
3.3.	Contribution of risk criteria to the estimated disease risk score amongst 120 producers in the New Zealand commercial poultry industry including 33 layer enterprises (including mixed pullet and layer operations), 57 broiler enterprises, 24 breeder enterprises and 6 other poultry enterprises (including duck, turkey and pullet operations). A risk score of zero indicates a minimal risk of disease introduction or spread where a score of 3 indicates a maximum level risk	p.99
3.4.	Level of concern over poultry health issues amongst 120 producers in the New Zealand commercial poultry industry including 33 layer enterprises (including mixed pullet and layer operations), 57 broiler enterprises, 24 breeder enterprises and 6 other poultry enterprises (including duck, turkey or pullet operations)	p.105
4.1.	Network analysis glossary of terms used to describe the trader networks and spatial networks representing poultry auction sales occurring through the TradeMe® website in New Zealand	p.131
4.2.	Summary network statistics for the New Zealand commercial poultry network showing the yearly on- and off-farms	p.137

movements relating to feed, live birds and hatching eggs, poultry waste and litter and, poultry products

- 4.3.** Frequency of poultry trades occurring through the online auction website TradeMe® in New Zealand stratified by the type of bird (n = 3) and year of trade **p.140**
- 4.4.** Estimated lower bounds on the number of poultry traded through the online auction website TradeMe® in New Zealand stratified by the type of bird (n = 3) and year of trade **p.140**
- 4.5.** Number of uniquely identified poultry traders utilizing the online auction website TradeMe® in New Zealand stratified the type of bird (n = 3) and year of trade **p.140**
- 4.6.** Summary network statistics for the trader network showing poultry trades occurring through the online auction website TradeMe® in New Zealand from 01st January 2018 to 31st December 2018 **p.143**
- 4.7.** Summary network statistics for the spatial network showing poultry trades occurring through the online auction website TradeMe® in New Zealand from 01st January 2018 to 31st December 2018 **p.144**
- 5.1.** The *C. jejuni* reference sequences and the corresponding NCBI references used to determine the likelihood of three plasmids: *pTet*, *pVir* and CP006703, and seven *C. jejuni* integrated elements (*CJIE1-4* and *CJIE1v*) being located in the *C. jejuni* ST-6964 genomes reported in this study (n = 167), using a method previously described by Llarena *et al.* 2016 **p.179**
- 5.2.** Network statistics describing both the empirical network (*i.e.*, the network constructed from 16 survey responses reporting all on- and off-farm movements over a one-year period) and the imputed network (*i.e.*, the network constructed using all 30 farms from which isolates were sampled, with missing edges imputed based on expert opinion) constructed from all on- and off-farm movements relating to (i) feed, (ii) live birds and hatching eggs or (iii) waste and litter within the New Zealand commercial poultry industry. Network metrics include the “Degree”, indicating the total number of on- and **p.193**

off-farm movements on a single farm in the network and, “Betweenness” indicating the frequency a farm is on the shortest path between any two other farms in the network

- 5.3.** Spearman’s rank matrix correlation (*rho*) between each model matrix and (i) the allelic dissimilarity matrix between 167 *C. jejuni* isolates and (ii) the residual allelic dissimilarity matrix after fitting the ANOVA factors of parent company ($n = 3$) and farm ($n = 30$) nested within parent company, with *p*-values obtained using 9999 permutations **p.194**

- 5.4.** Individual distance-based redundancy analysis (dbRDA) models to explain variation in allelic dissimilarities among 167 *C. jejuni* isolates in response to each of two ANOVA factors (parent company or farm nested in parent company) or sets of regression coordinates corresponding to geographic position (Euclidean distance or road distance) or the network model matrices of interest (feed, live birds or waste), with *p*-values for each of these separate marginal tests obtained using 9999 unrestricted permutations. “Prop” gives the proportion of the total variation explained whilst “*df*” gives the numerator (regression) and denominator (residual) degrees of freedom for the test. Models have been presented in order of decreasing R^2 values **p.196**

- 5.5.** Distance-based redundancy analysis (dbRDA) to explain variation in allelic dissimilarities among 167 *C. jejuni* isolates in response to the factors and sets of regressors listed in Table 5.4, but here conditional tests were done in a sequential stepwise fashion under forward selection based on R^2 . Each test used 9999 permutations of residuals under a reduced model. “Prop” gives the proportion of additional variation explained by adding that set of variables to the model, “Cumul” tracks the cumulative explained variation with each added step, and “*df*” provides the regression and residual degrees of freedom. **p.197**

- 5.6.** Top four dbRDA models obtained on the basis of the multivariate analogue to the Akaike Information Criterion (AIC) (see text) to explain variation in allelic dissimilarities among 167 *C. jejuni* isolates. Two ANOVA factors (parent **p.197**

company and farms nested in parent company) were included in all potential models *a priori*.

- 6.1.** Network analysis glossary of terms used to describe the contact network graphs representing the movement of goods and services between commercial poultry producers in New Zealand **p.226**

- 6.2.** Summary farm demographics across 156 broiler farms registered in the PIANZ-EPF database, of which 63.5% (99/156) have been inferred by randomly sampling from a kernel density estimate derived from a histogram produced using survey responses from 36.5% (57/156) of the broiler farms **p.235**

- 6.3.** Summary network measures for the inferred network graphs representing the yearly on- and off-farms movements relating to feed, live birds and hatching eggs and, poultry waste and litter between all 156 commercial broiler producers in New Zealand registered on the PIANZ-EPF database **p.236**

- 7.1.** Isolate-level (n = 57), farm-level (n = 17) and cow-level (n = 52) prevalence of eight *Staphylococcus aureus* whole-genome multi-locus sequence types (ST) isolated from the Waikato region of New Zealand North Island. The distribution of *S. aureus* sequence types among isolates show if the isolate was collected from a cow presenting with either clinical (n = 19) or sub-clinical (n = 38) mastitis **p.275**

- 7.2.** Drug classes influenced by the 14 resistance genes identified in the 57 *Staphylococcus aureus* isolates and their common uses in the New Zealand dairy industry. Resistance genes have been grouped by their gene family **p.278**

- 7.3.** Scoary result summary showing the genes found to be significantly associated with *Staphylococcus aureus* sequence type (ST)-1 and the Gene Ontology (GO) terms indicating either the biological processes or molecular functions associated with the gene products. Significant genes were those that had an odds ratio greater than 1, a specificity greater than 95%, a Benjamini-Hochberg corrected *p-value* below **p.281**

0.05 and were not annotated as either a “hypothetical” or “putative” protein

- 7.4. The presence-absence of 14 resistance genes and the antimicrobial sensitivity across 50 *S. aureus* isolates. Antimicrobial tested zone diffusion include penicillin (Pen), novobiocin (Nov), ceftiofur (Cef), tetracycline (Tet), ceftiofur (XNL), and oxacillin (OXA); with isolates being declared sensitive (S) or resistant (R)(including those also identified as intermediate), based on CLSI recommendations **p.284**

- 7.5. The dominant *Staphylococcus aureus* 7 gene multilocus sequence types (MLST) observed in our study looking at 57 *S. aureus* isolates sampled from bovine milk samples, and previous studies conducted by Grinberg *et al.* (2017) (Study 1) including 59 *S. aureus* isolates sampled from humans (n = 34) and household pets (cats = 6, dogs = 19) from across New Zealand and Heffernan *et al.* (2015) (Study 2) including 1255 methicillin-susceptible *S. aureus* (MSSA) isolates sampled from humans from across New Zealand. Note the MLST for five of the isolates in the current study (5/57, 8.8%) could be detected as well as three in study 1 (3/59, 5.1%). For study 2 only the seven dominant MSSA clones have been identified therefore, the column percentages do not equal 100% **p.288**

- 7.6. Summary network statistic for all on- and off-farm live animal movements recorded in the Livestock Improvement Corporation (LIC) MINDA database over a 26-year period from 1985-2011 **p.293**

- 7.7. Multivariate Permutational Multivariate Analysis of Variance (PERMANOVA) model comparing the *p*-dissimilarity measure between the core genome single nucleotide polymorphism (SNP) profiles of 57 *Staphylococcus aureus* isolates by network community (n = 6) and farm (nested within community) (n = 17) with *p-values* for each test obtained using 9999 unrestricted permutations. “SS” provides the sum of squares, “MS” the mean squares, and “*df*” the degrees of freedom for each test **p.295**

“And now that you don’t have to be perfect, you can be good.”

- John Steinbeck, 1952

CHAPTER 1

Introduction

1.1. Infectious disease dynamics in a connected world

A report published by The National Academy of Medicine, formerly known as the Institute of Medicine in the United States, in 1992 identified six key factors as the most likely driving forces behind future trends in emerging and re-emerging infectious diseases. These factors were (i) human demographics and behaviour, (ii) technology and industry, (iii) economic development and land use, (iv) international travel and commerce, (v) the breakdown of public health measures and, (vi) microbial adaptation and change (Institute of Medicine, 1992). Since then, there have been many additional factors proposed in relation to the global trends of infectious diseases and our capacity to effectively prevent, control and treat these diseases (Jones *et al.* 2008). Central to many of these discussions are the effects of population growth and increased global connectivity (Wilson, 1995; Institute of Medicine, 2010). Over the past century, both human and animal movement patterns have grown in distance, volume and speed; blurring the geographical boundaries for infectious diseases (Tatem *et al.* 2006) and as of yet, the full consequences of this increased connectivity on the spread, persistence, risk and control of diseases are just beginning to be understood.

In order to continue the fight against infectious diseases, it is clear that a greater understanding is needed into how population contact patterns, particularly those occurring at the animal-human-ecosystem interface, may be contributing to the changing patterns in disease emergence and spread (Lloyd-Smith *et al.* 2009; Hassell *et al.* 2017). Traditionally in epidemiological studies, mathematical compartmental model, such as a Susceptible–Infectious–Removed (SIR) model, have been used to describe the transmission dynamics of a disease within a population, with the earliest model described in work by Daniel Bernoulli (1700– 1782) on the inoculation of smallpox (Bernoulli, 1760). However, in recent years, a growing number of studies are turning to network-based approaches in order to overcome the simplifying assumption of homogeneous-mixing that is inherent to many compartmental models (Keeling *et al.* 2008). This

assumption, that all individuals within the population mix both uniformly and randomly, overlooks the influence of the population contact patterns on disease dynamics (Bansal *et al.* 2007) however, in many of these homogeneous mixing compartmental models, there is just not enough information known to inform the underlying population contact structure. Difficulties in accurately capturing population contact patterns, particularly within wildlife and livestock populations, is one of the major limitations to many of the network-based approaches that exist in the literature (Craft, 2015) and is major driver behind many new methods that aim at integrating multiple complementary data sources such as network data and pathogen sequence data.

1.2. Host contact networks and pathogen phylogenetics

Network-based approaches are grounded in graph theory; the basic concept of which was introduced back in the late-18th century by Leonhard Euler (1707-1783) (Euler, 1995). Since then, the development of graph theory has provided an array of quantitative tools for describing networks, many of which have been used across a number of research disciplines including sociology, epidemiology, psychology, computer science and economics (Otte and Rousseau, 2002; Borgatti *et al.* 2009; Scott and Carrington, 2011). Network graphs can be constructed from a set of elements, often referred to as nodes, vertices or actors, which represent the unit of interest, and edges or contacts to show the relationships between them. For example, in infectious disease studies nodes can represent individuals (*i.e.*, a human or an animal), or any larger epidemiological group (*i.e.*, a hospital or a farm), with nodes connected via different types of contacts that are known to be pathways for disease transmission. In systems with relatively complete information on the contact network nodes and edges, it may be possible to identify who-infected-whom as all transmission events will be captured by a contact in the network (Keeling and Eames, 2005; Danon *et al.* 2010). However, correctly identifying which contacts contribute towards the spread of disease is complicated by both incomplete

network data and multiple networks relevant to disease transmission existing within the host population (Eames *et al.* 2015).

Over the last few decades, the increasing availability of molecular sequence data has seen a rise in the number of studies using pathogen phylogenetics to strengthen the epidemiological inferences made from host contact networks (Pluciński *et al.* 2011; Vasylyeva *et al.* 2016; Gilbertson *et al.* 2018). This has also led to the rapid expansion of molecular epidemiological tools that aim at making inferences on disease dynamics based on observed genetic mutations between sampled pathogen sequences and attached epidemiological information. This basic concept sets the foundation for most molecular epidemiological studies (Hall, 1996) in addition to many other studies that identify with other research disciplines but similarly use approaches grounded in population genetics to infer disease transmission dynamics from the population demographics of the pathogen population. However, the development of epidemiological methods that can integrate such evolutionary dynamics are still in their infancy and it remains unclear the impact of phylogenetic complexities on the validity of these new approaches (O’Dea and Wilke, 2011; de Maio *et al.* 2016).

1.3. Research focus

Within New Zealand, there is a growing concern for the transmission of zoonotic pathogens from farmed animals to humans. Recent disease outbreaks have served as a reminder of the potential impacts of endemic diseases such as, the 2016 campylobacteriosis outbreak in Havelock North that resulted in an estimated 5,500 cases, 45 hospitalisations, and 4 deaths (Ministry of Health, 2017). In addition to public health consequences, the economy of New Zealand is highly dependent on the health of its livestock populations and recent disease incursions, such as *Mycoplasma bovis* (Roche, 2019), has further highlighted the potential impact of exotic diseases on not only the health and welfare of livestock but also the stability of New Zealand’s primary sector and

access to global trade. Therefore, in order to protect public health, ensure the wellbeing of livestock populations and, maintain a competitive edge in global trade markets, we must develop new approaches to minimise the impact of endemic diseases, reduce the transmission of zoonotic diseases and prevent the incursion of exotic diseases. However, many modern farming practices have created unique opportunities for disease to spread between animal populations through the movements of animals, vectors, personnel, and equipment. The continuing development of the urban landscape is also creating diverse wildlife–livestock–human interfaces that represents a critical point for the transmission and emergence of zoonotic diseases, although the interacting patterns occurring at these interfaces are not always clear (Karesh *et al.* 2012; Reperant *et al.* 2013). Only with a better understanding of how these contacts and behaviours shape pathogen evolution and transmission dynamics, will it be possible to design more cost-effective disease control strategies in often what is a resource-limited setting.

This thesis forms part of a larger network of studies funded through New Zealand’s Biological Heritage (NZBH) National Science Challenge in a project collectively entitled “Biosecurity Network Interventions”. Across all the studies, the overarching aim is to bring together multiple disciplinary perspectives in network modelling in order to identify opportunities for intervention and reduce the spread of pests, pathogens and weeds within four human-assisted networks across New Zealand these are; (i) the plant nursery network, (ii) the livestock transport network, (iii) the freshwater recreational user network and (iv) the natural area visitor network. The main focus of this thesis is looking at how livestock transport networks can be used to inform different stages of an epidemiological investigation with a particular interest in integrating pathogen whole-genome sequence data within a network analysis to help determine the contribution of different network pathways on pathogen population dynamics. This thesis describes two livestock networks, one constructed from the movement of goods and services within the commercial poultry industry and another from the movement of live-animals within the

commercial dairy cattle industry, in order to investigate the transmission and evolutionary dynamics of two pathogens; *Campylobacter jejuni* within the poultry network and *Staphylococcus aureus* within the dairy cattle network.

1.3.1. New Zealand's commercial poultry industry

The historic introduction of poultry into New Zealand is highly disputed with records dating back to the late-1700s with the arrival of Captain James Cook (1728-1779) (Wintle and Lepper, 2012). In the decades to come production remained largely subsistence with an estimated half of all New Zealand households keeping hens in their backyard to supply eggs (Binney *et al.* 2014). However, during the early 20th century several events resulted in a gradual shift from predominantly backyard production to small-scale production of both meat and eggs including increased recognition by the New Zealand Government for the nutritional value of eggs, increased demand for chicken by US hospitals in the Pacific for recovering soldiers following the Second World War, improved production techniques, the lifting of import restrictions from outside Australia, and the increasing popularity of chicken meat largely driven by rising public affluence, recognition of health benefits in comparison to red meat, increased popularity of international cuisines and convenience (Stafford, 2017).

Today modern poultry production in New Zealand focuses around three major species (chickens, turkeys and ducks) plus several minor species including geese, guinea fowl, quails, pheasants and ostriches. The commercial chicken industry is the largest of these consisting of approximately 119 million meat chickens raised annually (mainly Cobb and Ross), 3.5 million layer chickens, with a further 3 million replacements raised each year (mainly Shaver Browns and Hyline Whites), and 2.5 million meat and layer breeder birds (StatsNZ, 2018). The market is largely domestic, however, increasing amounts of meat and eggs are being exported due to international recognition of New Zealand's high poultry health status with New Zealand national flock remaining free from major avian

disease including both Newcastle disease and Highly Pathogenic Avian Influenza (Davidson, 2002; Cobb and Smith, 2013) and until recently, Infectious Bursal Disease (MPI, 2019a).

Industry structure varies by sector (Figure 1.1) with the broiler industry characterised by a small number of large vertically integrated companies such that, the four largest poultry-meat suppliers; that is, Tegel Foods, Inghams Enterprises, PH van den Brink, and Turks Poultry, supplies over 90% of the country's poultry meat. This structure supports the vertical integration of most broiler operations with individual producers relying on suppliers to control all stages from primary production and processing to distribution. In comparison, the layer industry is dominated by many smaller independent producers that rely on horizontal integration and depend almost entirely on a domestic market with many producers selling directly to the wholesalers within a single administrative boundary or marketing through co-operatives (Stafford, 2017).

1.3.2. Epidemiology of *Campylobacter jejuni* in New Zealand

Within the commercial poultry industry there are many concerns associated with production it may be argued that considering the high health status of New Zealand national flock and strict industry guidelines to ensure that both environmental contamination is minimised and animal welfare standards are maintained (Coriolis, 2017), that food safety remains one of the primary concerns for the industry. Within the broiler industry, this includes the control of pathogens responsible for major foodborne illnesses such as *Campylobacter*, *Salmonella* and *Listeria* whereas the primary concern for egg producers will be *Salmonella*. In particular, a major concern in New Zealand is *Campylobacter* with incidence rates of human Campylobacteriosis greatly exceeding that in comparison to other developed countries (Olson *et al.* 2008).

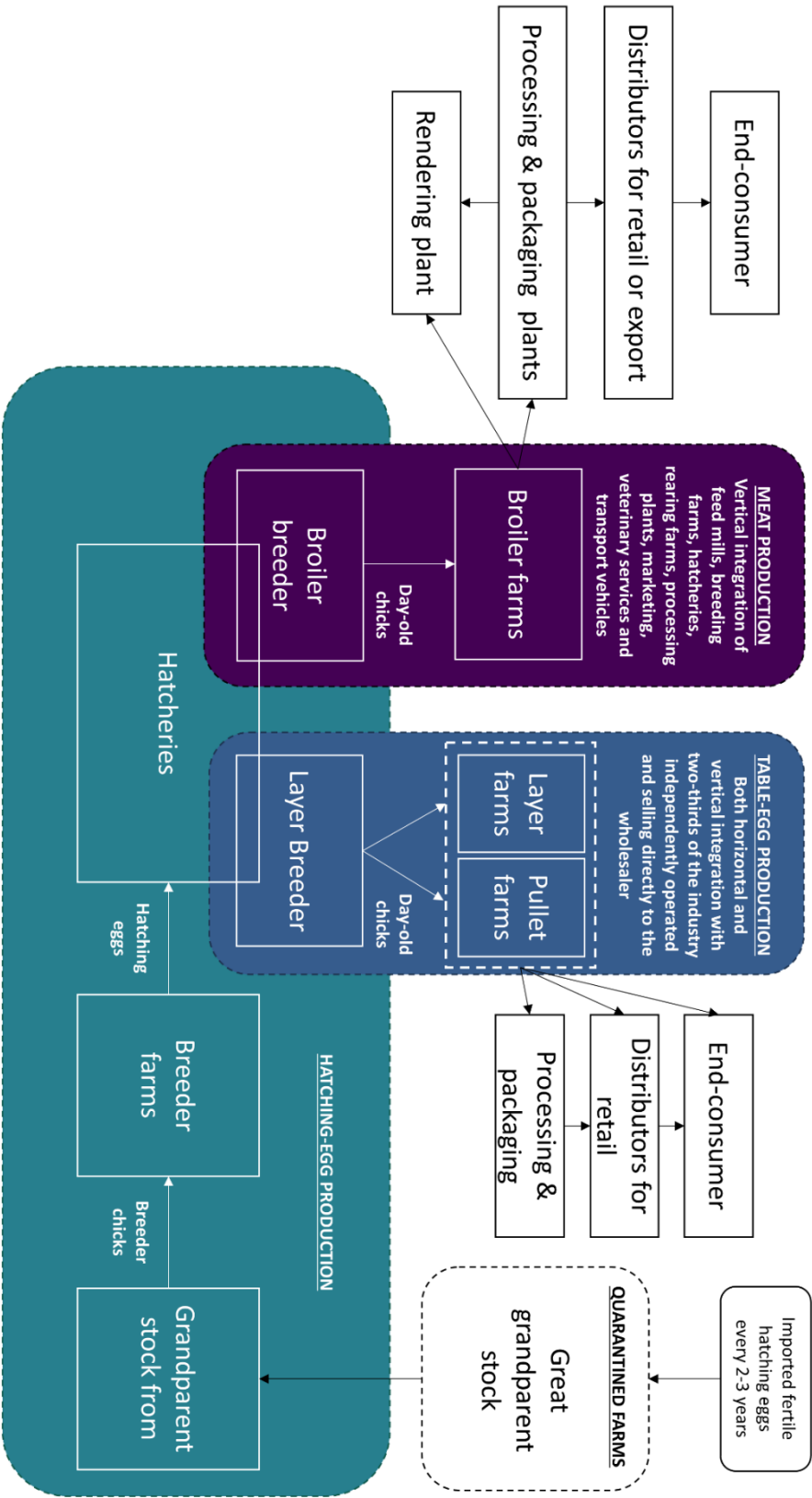


Figure 1.1. Schematic diagram showing the structure of the New Zealand poultry industry including breeder farms and hatcheries (green), broiler farms (purple), and layer farms (blue).

Campylobacteriosis is a foodborne disease caused by an infection from a number of Gram-negative bacterial species belonging to the genus *Campylobacter* (Moore *et al.* 2005). In New Zealand, the disease is typically attributed to infection with *C. jejuni* and to a lesser degree, *C. coli* and *C. fetus* (Müllner *et al.* 2010) with infection most often resulting in acute, self-limiting gastroenteritis. Common symptoms mostly include abdominal pain, diarrhoea, vomiting, nausea, and fever; however, occasionally severe and persistent gastroenteritis will result in hospitalisation with macrolides and fluoroquinolones recommended as the first-choice antimicrobials for treating the more severe cases or immunocompromised individuals. A further 5–10% of cases are also associated with post-infectious extraintestinal complications such as meningitis, carditis, pancreatitis, urinary tract infections, reactive arthritis, and on rare occasions, the neurological syndrome Guillain-Barre (WHO, 2012).

Many animals act as natural hosts for *Campylobacter* spp. with many showing little to no sign of carriage including most species of domestic animals such as cattle, sheep, poultry, pigs, dogs and cats (Blaser *et al.* 1984). *Campylobacter* can also survive in host faecal matter and although recent genotyping has demonstrated some degree of host association between *Campylobacter* spp. it is known that environmental contamination can also act as an indirect route of transmission between different host species as well as having the potential to contaminate other food sources, such as fruits and vegetables, through contact with contaminated soil or water sources (Whiley *et al.* 2013). In New Zealand, source attribution studies have indicated that a high proportion of campylobacteriosis cases are due to the ingestion of contaminated food, most frequently undercooked poultry (Müllner *et al.* 2009). This prompted the New Zealand Food Safety Authority into launching New Zealand's first *Campylobacter* risk management strategy back in 2006, focusing on reducing the contamination of chicken meat with the implementation of both mandatory and voluntary control strategies along the poultry supply chain (NZFSA, 2008). As a result of control efforts, the number of notified cases declined from 379 cases

per 100,000 in the population in 2006 to 125 cases per 100,000 in the population in 2017 (Sears *et al.* 2011).

However, in May 2014, a previously unreported *C. jejuni* sequence type (ST) 6964 was found after routine sampling at a sentinel surveillance site in the Manawatu region of New Zealand (Bolwell *et al.* 2015). Following the emergence of *C. jejuni* ST-6964, several cross-sectional studies were conducted in both poultry and humans in order to investigate possible changes in the resistance patterns of *C. jejuni*. Results from these studies found 30% of human *C. jejuni* isolates to be resistant to fluoroquinolones, 77 % of which were also resistant to tetracycline, and 37% of poultry *C. jejuni* isolates to be resistant to both ciprofloxacin and tetracycline (Muellner *et al.* 2016). In addition to the significant increase in resistance, this ST-6964 was also found to be widely distributed across all four major poultry producers (Muellner *et al.* 2016) despite the industry being vertically integrated with little contact between producers belonging to different suppliers. This suggested a major epidemiological shift as previously the dominant ST in any given year was typically associated with individual suppliers (Müllner *et al.* 2010). In the years to follow many human cases of campylobacteriosis associated with ST-6964 have been identified in outbreaks across New Zealand however, the origin and transmission dynamics of this ST remain unclear.

1.3.3. New Zealand's dairy cattle industry

The New Zealand dairy industry has a history dating back to the early-1800s when the first shorthorn cattle were imported into New Zealand by the Reverend Samuel Marsden (1765-1838) (Peden, 2008). In the following years, more cattle breeds were imported into New Zealand establishing the parent breeds of the modern national herd including Ayrshire, Jersey and Friesian cattle. Dairy export was limited during this time with only small amounts of butter and cheese being exported to Australia up until the mid-1880s when refrigeration was developed. Further development of new technologies, such as the

mechanical milking machine in 1918, and improvements in genetics, pastures and feeding resulted in almost a continuous increase in both the total number of cows and the number of cows per herd ever since (Stafford, 2017). Recent results from the 2017 agricultural survey now estimate the total number of dairy cattle to have reached over 6.5 million cows of which the dominant breed is Holstein Friesian x Jersey crosses, otherwise known as KiwiCross cows (Morris, 2013). In New Zealand, dairy production is largely focused on a pasture-based system with seasonal calving taking place in the spring (July-October) (Figure 1.2). This varies greatly in comparison to countries in the northern hemisphere where there is a greater reliance on indoor housing systems and year-round calving. The industry is highly integrated with processors encompassing production, manufacturing and marketing. This structure was established with The Dairy Board back in 1923, which was responsible for the marketing and sale of all milk products until 2001 when the two largest dairy processors merged with The Dairy Board to form Fonterra Cooperative Group Ltd. who now controls approximately 85% of the raw milk produced in the country (Stringleman and Scrimgeour, 2008).

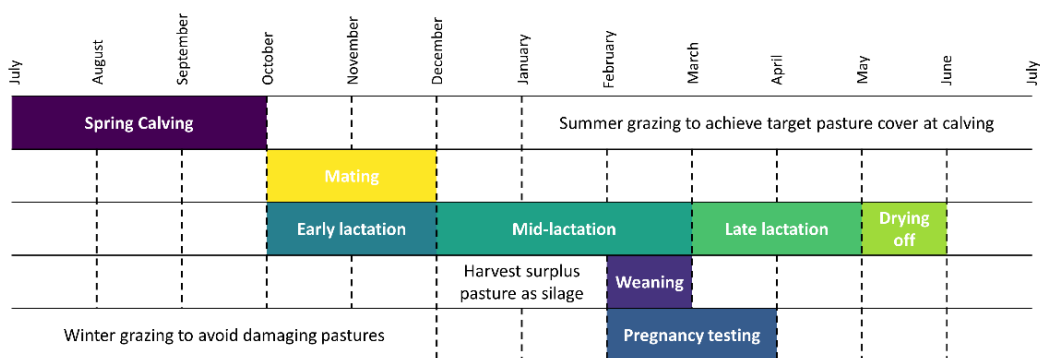


Figure 1.2. A typical annual production cycle for New Zealand dairy farms using a pasture-based system with seasonal calving taking place in the spring (purple; July-October), followed by mating (yellow; October-December), lactation and drying off (green; October-June), and weaning and pregnancy testing (blue; February-April).

The growth of New Zealand's dairy industry is reflected in the contribution the dairy sector makes to the New Zealand economy. In 2018 it was calculated that the dairy sector

contributed approximately \$8.7 billion New Zealand dollars (NZD) to New Zealand's total gross domestic product (GDP). This is largely dependent on the export market with over 95% of milk exported, accounting for approximately 30% of all dairy products on the global market (MPI, 2019b). Continued access to the global market was one incentive that led to the development of the National Animal Identification and Tracing System (NAIT) in 2009 at a time when the World Animal Health Organisation (OIE) was actively encouraging the standardisation of animal identification and tracking systems in order to improve the management of livestock diseases and provide assurance within the global market of food safety, traceability and quality. The current NAIT scheme aims to link producers, property and livestock using radio frequency identification device (RFID) ear tags.

The strengths and weaknesses of NAIT in providing traceability of individual animals was tested with a 2017 outbreak of *M. bovis*, a bacterial disease that can lead to serious conditions in cattle impacting both animal welfare and productivity. In July 2017, the disease was detected on a South Island dairy farm and subsequent contact tracing, surveillance and testing have resulted in a further 51 properties having undergone phased eradication (MPI, 2018). During this time NAIT was highly criticized however, it emphasised the important contribution of live animal movements towards infectious disease dynamics and the significance of being able to capture livestock contact networks in order to rapidly be able to respond to a disease outbreak. Despite this, the use of NAIT data in infectious disease research is still limited particularly for endemic diseases. In addition to NAIT, the Livestock Improvement Corporation (LIC), a multinational farmer-owned co-operative, provides members with a computerised herd management software called MINDA which can provide information on both the movement of adult lactating dairy cows as well as basic production data making it ideal for modelling the spread of contagious pathogens within the dairy industry such as those pathogens responsible for mastitis.

1.3.4 Epidemiology of *Staphylococcus aureus* in New Zealand

New Zealand is free from many infectious diseases that are known to cause significant losses in cattle such as Vesicular Stomatitis Virus (VSV), Enzootic Bovine Leucosis (EBL) and brucellosis (Vermunt, 2000; Davidson, 2002; Seleem *et al.* 2010) with eradication programs in place for many others including Bovine Viral Diarrhoea Virus (BVDV) (Han *et al.* 2018). However, one disease well recognised as imposing major production costs on the dairy industry is bovine mastitis with annual costs in New Zealand estimated to be over \$180 NZD (Petrovski, 2007). Bovine mastitis is an inflammation of the mammary gland in response to physical trauma or a pathogenic infection. Infection can result in either sub-clinical or clinical disease with clinical mastitis presenting a number of symptoms including udder abnormalities (such as swelling, heat, hardness, redness, or pain) and milk defects (such as a reduction in yield, a watery appearance, a change in milk colour, flakes, clots, or pus). The severity of symptoms is influenced by many factors including several environmental variables such as humidity and temperature as well as the nutritional or immune status of the cow, however, one of the biggest influencing factors is the responsible pathogen (Eberhart, 1986; Bogni *et al.* 2011; Gomes *et al.* 2016).

Bovine mastitis can be caused by over 137 different organisms including bacteria, yeasts and algae (Watts, 1988). The relative importance of these different pathogens is largely country dependent however, more than 90% of all new intra-mammary infections in dairy cattle are caused by a small number of pathogenic bacteria namely *Escherichia coli*, *Streptococcus dysgalactiae*, *Streptococcus uberis*, *Streptococcus agalactiae* and *Staphylococcus aureus* (Bradley, 2002). Traditionally, these mastitis-causing pathogens can be classified as either contagious or environmental pathogens based on epidemiological observations such that contagious pathogens spread from cow-to-cow usually during the milking process whereas environmental pathogens spread from a contaminated environment such as bedding, soil or manure (Watts, 1988). For some pathogens, the distinction between contagious and environmental is not clear. For

example, some *Streptococcus spp.* that have previously been classed as environmental pathogens are now also considered contagious in some circumstances however, historically the major contagious pathogens are considered to include *S. agalactiae*, *S. aureus* and *Mycoplasma spp.* and the most common environmental pathogens include *E. coli*, *S. uberis*, *S. dysgalactiae* and many *Klebsiella spp.*

In New Zealand the epidemiology of the predominant mastitis-causing pathogens varies greatly in comparison to countries in the northern hemisphere where there is a greater reliance on indoor housing systems and year-round calving practices (McDougall, 2002; Heffernan *et al.* 2015). Within the last five decades, New Zealand has seen a notable change in the contribution of the major pathogens responsible for mastitis in dairy cows with major changes first noted following the introduction of the mastitis control strategy in the early 1990s. The Seasonal Approach to Managing Mastitis (SAMM) plan was introduced by The National Mastitis Advisory Committee in conjunction with an industry initiative to reduce the incidence of mastitis through the selective breeding of dairy cows. The current strategy for mastitis control (SmartSAMM) builds on the original SAMM plan with an additional drive for antibiotic dry-cow therapy and the treatment of intra-mammary infection during drying off (Lacy-Hulbert *et al.* 2011). The successful uptake of these programs by New Zealand farmers has resulted in an epidemiological transition with a decreasing prevalence of contagious pathogens, such as *S. agalactiae*, in comparison to environmental pathogens which are now on the rise (McDougall *et al.* 2007; Petrovski *et al.* 2011). Despite this transition, a significant proportion of the total cost of mastitis for New Zealand dairy farmers is attributed to infections associated with the contagious pathogen *S. aureus* as it is most commonly associated with chronic subclinical mastitis that is both hard to detect and hard to treat successfully.

Staphylococcus is a genus of Gram-positive cocci belonging to the family *Staphylococcaceae*. Infections with *S. aureus* may result in clinical mastitis, especially after

calving, however, the infection is usually subclinical, causing elevated somatic cell counts but no detectable changes in milk or the udder (Bonsaglia *et al.* 2018). *S. aureus* infections are notoriously difficult to treat with antibiotics alone, therefore, based on the costs and benefits of treatment, it is commonly advised that infected cow be removed from the herd (Sandholm *et al.* 1990). The difficulty with antimicrobial therapy is thought to be due to numerous biofilm-associated genes (BAGs) that are known to play a role in mechanisms that assist the pathogen in evading the host's immune response and acquiring multi-drug resistance (Kot *et al.* 2018). Antimicrobial resistant *S. aureus* has been widely documented in many countries creating a growing concern over the use of antimicrobials for the treatment and control of mastitis. This is not only due to the general concern of using antibiotics in food-producing livestock, but also more specifically to the increasing presence of methicillin-resistant *S. aureus* (MRSA) infection in dairy cattle alongside evidence for livestock-associated methicillin-resistant *S. aureus* (LA-MRSA) infections in humans creating a huge concern for both livestock and public health (Wulf and Voss, 2008; Hillerton and Allison, 2015).

Since the first LA-MRSA was described in Belgium in 1972, deriving from a case of bovine mastitis (Deriese *et al.* 1972), there has been an influx of studies looking at the emergence, evolution and dissemination of LA-MRSA with particular attention given to the most widely spread LA-MRSA clonal complex (CC) 398. The first communications on LA-MRSA CC398 originated from nasal swabs isolated from breeding pigs, however, since then it has been reported in veal calves, poultry, dairy cattle, goats, cats, dogs, mice, rats and horses (Price *et al.* 2012; Mohammed and Nigatu, 2015). The mechanisms of methicillin resistance have been well documented with MRSA evolving from methicillin-susceptible *S. aureus* (MSSA) by the acquisition of SSCmec elements containing a *mec* gene, most commonly *mecA* but also *mecC* (Aires de Sousa, 2017). These *mec* genes code for an additional penicillin-binding protein that has low affinity for β -lactam antibiotics, therefore, mediating resistance to nearly all compounds from this antibiotic class which

are still used frequently to treat mastitis. Recent studies in New Zealand looking at the susceptibility of *S. aureus* have found increases in the prevalence of both fusidic acid resistance and mupirocin resistance (McDougall *et al.* 2014; Petrovski *et al.* 2015) however, despite these findings, the antimicrobial susceptibility patterns of mastitis-causing pathogens in New Zealand and their threat to the future control of bovine mastitis and public health remains unclear.

1.4. Thesis structure

This thesis begins with a literature review (Chapter 2) that aims at expanding the introduction by presenting and critiquing some of the existing epidemiological tools in the literature that utilises either host contact networks or pathogen sequence data to help infer disease transmission dynamics, as well as, outlining some of the basic principles integral to both network analysis and phylogenetics and discussing both the limitations and future challenges that need to be considered in order to successfully integrate the two data sources. This review is then followed by a series of five research chapters (Chapters 3-7) with each chapter having different research objectives that align with the research focus. Repetition between the chapters has been kept to a minimum however, all chapters have been prepared as a manuscript intended for publication in a peer-reviewed journal, resulting in a small amount of repetition throughout the thesis.

Chapters 3-5 focus largely on New Zealand's commercial poultry industry with Chapter 3 reporting the results from a cross-sectional survey that was conducted across all commercial poultry operations in 2016. Chapter 4 expands on the survey results presented in the previous chapter to reconstruct contact networks from the reported on- and off-farm movements of goods and services and investigate the disease transmission risk within the commercial poultry industry from both the sale of backyard poultry and the migration of wild birds. Chapter 5 describes the genetic population structure of 167 *C. jejuni* isolates from a sample of commercial poultry farms whilst further using the

contact networks to investigate the relative importance of different contact types as determinants of the pairwise genetic relatedness between the isolates. Chapter 6 conducts a similar analysis to that in the previous chapter but instead uses simulated *C. jejuni* sequence data generated from a network simulation model in order to validate the methods used in Chapter 5. Chapter 7 then moves on to describe the genetic population structure of 57 bovine-derived *S. aureus* isolates from a sample of New Zealand dairy herds, determining their relationship with other *S. aureus* isolates collected previously from domesticated pets and humans both within New Zealand and internationally in order to demonstrate how a range of phylogenetic tools and network analysis can be used to further investigate transmission dynamics at a range of geographical scales. Following this research chapter, there is a general discussion (Chapter 8) that summarises the main research findings from Chapters 3-7 whilst giving further consideration to the limitations present in each study, the potential application of results in the real world and how they may be used to guide future research.

1.5. References

- Bansal S, Grenfell BT, Meyers LA. When individual behaviour matters: homogeneous and network models in epidemiology. *J. R. Soc. Interface* 4, 879–891, 2007
- Bernoulli D. *Reflexions sur les avantages de l'inoculation*. Mercure de France, Paris, France, 1760. English translation by R. Pulskamp (2009). Available at http://www.cs.xu.edu/math/Sources/DanBernoulli/1760_reflections_inoculation.pdf. Department of Mathematics and Computer Science, Xavier University, Cincinnati
- Binney BM, Biggs PJ, Carter PE, Holland BR, Holland BM, French NP. Quantification of historical livestock importation into New Zealand 1860-1979. *N. Z. Vet. J.* 62, 309–314, 2014
- Blaser MJ, Taylor DN, Feldman RA. *Epidemiology of Campylobacter infections*. In: *Campylobacter Infection in Man and Animals*. Pp 142–162, CRC Press, Florida, United States, 1984
- Bogni CI, Odierno LM, Raspanti CG, Giraudo JA, Larriestra A, Lasagno MC, Ferrari MH, Ducros E, Frigerio C, Bettera SG, Frola ID, Dieser SA, Vissio C. *War against mastitis: Current concepts on controlling bovine mastitis pathogens*. In: *Science against microbial pathogens: communicating current research and technological advances*, Formatex Research Center, Badajoz, Spain, 2011
- Bonsaglia ECR, Silva NCC, Rossi BF, Camargo CH, Dantas STA, Langoni H, Guimarães FF, Lima FS, Fitzgerald JR, Fernandes A, Rall VLM. Molecular epidemiology of methicillin-susceptible *Staphylococcus aureus* (MSSA) isolated from milk of cows with subclinical mastitis. *Microb. Pathogen.* 124, 130–135, 2018
- Borgatti SP, Mehra A, Brass DJ, Labianca G. Network analysis in the social sciences. *science* 323, 892–895, 2009
- Bradley AJ. Bovine mastitis: an evolving disease. *Vet. J.* 164, 116–128, 2002
- Cobb S, Smith H. Import risk analysis: Chicken and duck meat for human consumption. Available at: <https://www.mpi.govt.nz/dmsdocument/2815>. Accessed 13 September 2019. Ministry for Primary Industries, Wellington, New Zealand, 2013
- Coriolis. *The investor's guide to the New Zealand meat industry 2017*. Ministry of Business, Innovation and Employment (MBIE), Auckland, New Zealand, 2017

- Craft ME.** Infectious disease transmission and contact networks in wildlife and livestock. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370, e20140107, 2015
- Danon L, Ford AP, House T, Jewell CP, Keeling MJ, Roberts GO, Ross JV, Vernon MC.** Networks and the epidemiology of infectious disease. *Interdis. Perspec. Infec. Dis.* 2011, 1–28, 2011
- Davidson RM.** Control and eradication of animal diseases in New Zealand. *N. Z. Vet. J.* 50, 6–12, 2002
- de Maio ND, Wu C-H, Wilson DJ.** SCOTTI: efficient reconstruction of transmission within outbreaks with the structured coalescent. *PLoS Comput. Biol.* 12, e1005130, 2016
- Devriese L, R Van Damme L, Fameree L.** Methicillin (Cloxacillin)-Resistant *Staphylococcus aureus* strains Isolated from Bovine Mastitis Cases. *J Vet. Med.* 19, 598–605, 1972
- Eames K, Bansal S, Frost S, Riley S.** Six challenges in measuring contact networks for use in modelling. *Epidemics* 10, 72–77, 2015
- Eberhart RJ.** Management of dry cows to reduce mastitis. *J. Dairy Sci.* 69, 1721–1732, 1986
- Euler L.** *From the problem of the seven bridges of Königsberg.* In: Classics of Mathematics, Englewood Cliffs, New Jersey, US, 1995
- Gilbertson MLJ, Fountain-Jones NM, Craft ME.** Incorporating genomic methods into contact networks to reveal new insights into animal behaviour and infectious disease dynamics. *Behaviour* 155, 759–791, 2018
- Gomes F, Saavedra MJ, Henriques M.** Bovine mastitis disease pathogenicity: evidence of the potential role of microbial biofilms. *Pathog. Dis.* 74, ftw006, 2016
- Hall A.** What is molecular epidemiology? *Trop. Med. Int. Health* 1, 407–408, 1996
- Han J-H, Weir AM, Weston JF, Heuer C, Gates MC.** Elimination of bovine viral diarrhoea virus in New Zealand: a review of research progress and future directions. *N. Z. Vet. J.* 66, 273–280, 2018
- Hassell JM, Begon M, Ward MJ, Fèvre EM.** Urbanization and disease emergence: dynamics at the wildlife–livestock–human interface. *Trends Ecol. Evol.* 32, 55–67, 2017

- Heffernan H, Bakker S, Woodhouse R, Dyet K, Williamson DA.** *Demographics, antimicrobial susceptibility and molecular epidemiology of Staphylococcus aureus in New Zealand, 2014*. Institute of Environmental Science and Research Limited, Wellington, New Zealand, 2015
- Institute of Medicine.** Emerging infections: microbial threats to health in the United States. Available at: <https://doi.org/10.17226/2008>. Accessed 2 October 2019. The National Academies Press, Washington, DC, 1992
- Institute of Medicine.** Infectious disease movement in a borderless world: Workshop Summary. Available at: <https://doi.org/10.17226/12758>. Accessed 2 October 2019. National Academies Press, Washington, DC, 2010
- Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P.** Global trends in emerging infectious diseases. *Nature* 451, 990–993, 2008
- Karesh WB, Dobson A, Lloyd-Smith JO, Lubroth J, Dixon MA, Bennett M, Aldrich S, Harrington T, Formenty P, Loh EH, Machalaba CC, Thomas MJ, Heymann DL.** Ecology of zoonoses: natural and unnatural histories. *The Lancet* 380, 1936–1945, 2012
- Keeling MJ, Eames KTD.** Networks and epidemic models. *J. R. Soc. Interface* 2, 295–307, 2005
- Keeling MJ, Rohani P, Pourbohloul B.** Modeling infectious diseases in humans and animals. *Clin. Infect. Dis.* 47, 864–865, 2008
- Kot B, Sytykiewicz H, Sprawka I.** Expression of the biofilm-associated genes in methicillin-resistant *Staphylococcus aureus* in biofilm and planktonic conditions. *Int. J. Mol. Sci.* 19, 2018
- Lacy-Hulbert J, Blackwell M, McDougall S.** *SmartSAMB: The smart approach to minimising mastitis*. Proceedings of the Society of Dairy Cattle Veterinarians of the NZVA Annual Conference, Wellington, New Zealand, 2011
- Lloyd-Smith JO, George D, Pepin KM, Pitzer VE, Pulliam JRC, Dobson AP, Hudson PJ, Grenfell BT.** Epidemic dynamics at the human-animal interface. *Science* 326, 1362–1367, 2009
- McDougall S.** Bovine mastitis: epidemiology, treatment and control. *N. Z. Vet. J.* 50, 81–84, 2002

- McDougall S, Arthur DG, Bryan MA, Vermunt JJ, Weir AM.** Clinical and bacteriological response to treatment of clinical mastitis with one of three intramammary antibiotics. *N. Z. Vet. J.* 55, 161–170, 2007
- McDougall S, Hussein H, Petrovski K.** Antimicrobial resistance in *Staphylococcus aureus*, *Streptococcus uberis* and *Streptococcus dysgalactiae* from dairy cows with mastitis. *N. Z. Vet. J.* 62, 68–76, 2014
- Morris ST.** *Sheep and beef cattle production systems*. Manaaki Whenua Press, Lincoln, New Zealand, 2013
- Ministry for Primary Industries (MPI).** *Mycoplasma bovis - current situation. Documents, guidance, and fact sheets - Mycoplasma bovis resources, advice, fact sheets, and guidance for farmers, industry professionals, and the public.* Available at: <https://www.biosecurity.govt.nz/dmsdocument/32500-current-situation-infographic-7-dec-2018>. Accessed 2 October 2019. Ministry for Primary Industries, Wellington, New Zealand, 2018
- Ministry for Primary Industries (MPI).** *Protection and responses: Infectious bursal disease virus (IBDV).* Available at: <https://www.biosecurity.govt.nz/protection-and-response/responding/alerts/infectious-bursal-disease-virus-ibdvl/>. Accessed 22 October 2019. Ministry for Primary Industries, Wellington, New Zealand, 2019a.
- Ministry for Primary Industries (MPI).** *Situation and Outlook for Primary Industries March 2019.* Available at: <https://www.mpi.govt.nz/dmsdocument/33361/direct>. Accessed 22 October 2019. Ministry for Primary Industries, Wellington, New Zealand, 2019b
- Muellner P, Kells N, Campbell D.** *Risk profile the emergence of Campylobacter jejuni ST 6964 in poultry in New Zealand and its associated antimicrobial resistance.* MPI Technical Paper No: 2016/16, Wellington, New Zealand, 2016
- Müllner P, Spencer SEF, Wilson DJ, Jones G, Noble AD, Midwinter AC, Collins-Emerson JM, Carter P, Hathaway S, French NP.** Assigning the source of human campylobacteriosis in New Zealand: A comparative genetic and epidemiological approach. *Infect. Genet. Evol.* 9, 1311–1319, 2009
- New Zealand Food Safety Authority (NZFSA).** *Campylobacter risk management strategy.* Available at: <http://www.nzfsa.govt.nz/>. Accessed 12 February 2017. New Zealand Food Safety Authority, Wellington, New Zealand, 2008

- O'Dea EB, Wilke CO. Contact heterogeneity and phylodynamics: how contact networks shape parasite evolutionary trees. *Interdiscip. Perspect. Infect. Dis.* e238743, 2011
- Olson CK, Ethelberg S, Pelt W van, Tauxe RV. Epidemiology of *Campylobacter jejuni* infections in industrialized nations. In: *Campylobacter*. Pp 163–189. ASM Press, Washington, DC, 2008
- Otte E, Rousseau R. Social network analysis: a powerful strategy, also for the information sciences. *J. Inform. Sci.* 28, 441–453, 2002
- Peden R. Beef farming - 19th-century beef farming. Available at: <https://teara.govt.nz/en/beef-farming/page-2>. Accessed 2 October 2019. Te Ara - the Encyclopaedia of New Zealand, 2008
- Petrovski KR. *Bovine mastitis in New Zealand*. Master of Veterinary Science, Massey University, Palmerston North, 2007
- Petrovski KR, Grinberg A, Williamson NB, Abdalla ME, Lopez-Villalobos N, Parkinson TJ, Tucker IG, Rapnicki P. Susceptibility to antimicrobials of mastitis-causing *Staphylococcus aureus*, *Streptococcus uberis* and *Str. dysgalactiae* from New Zealand and the USA as assessed by the disk diffusion test. *Aust. Vet. J.* 93, 227–233, 2015
- Petrovski KR, Laven RA, Lopez-Villalobos N. A descriptive analysis of the antimicrobial susceptibility of mastitis-causing bacteria isolated from samples submitted to commercial diagnostic laboratories in New Zealand (2003-2006). *N. Z. Vet. J.* 59, 59–66, 2011
- Price LB, Stegger M, Hasman H, Aziz M, Larsen J, Andersen PS, Pearson T, Waters AE, Foster JT, Schupp J, Gillece J, Driebe E, Liu CM, Springer B, Zdovc I, Battisti A, Franco A, Żmudzki J, Schwarz S, Butaye P, Jouy E, Pomba C, Porrero MC, Ruimy R, Smith TC, Robinson DA, Weese JS, Arriola CS, Yu F, Laurent F, Keim P, Skov R, Aarestrup FM. *Staphylococcus aureus* CC398: host adaptation and emergence of methicillin resistance in livestock. *mBio* 3, e00305, 2012
- Reperant LA, Cornaglia G, Osterhaus ADME. The importance of understanding the human-animal interface: from early hominins to global citizens. *Curr. Top. Microbiol. Immunol.* 365, 49–81, 2013
- Roche J. *Mycoplasma bovis* casing and liaison backlog. Available at: <https://www.biosecurity.govt.nz/dmsdocument/35538-dr-john-roche-review>. Accessed 2 October 2019. Ministry for Primary Industries, Wellington, New Zealand, 2019

- Sandholm M, Kaartinen L, Pyörälä S.** Bovine mastitis - why does antibiotic therapy not always work? An overview. *J. Vet. Pharmacol. Therap.* 13, 248–260, 1990
- Scott J, Carrington PJ.** *The SAGE handbook of social network analysis*. SAGE Publication Inc., London, UK, 2011
- Sears A, Baker MG, Wilson N, Marshall J, Muellner P, Campbell DM, Lake RJ, French NP.** Marked Campylobacteriosis decline after interventions aimed at poultry, New Zealand. *Emerg. Infect. Dis.* 17, 1007–1015, 2011
- Seleem MN, Boyle SM, Sriranganathan N.** Brucellosis: A re-emerging zoonosis. *Vet. Microbiol.* 140, 392–398, 2010
- Stafford K.** *Livestock Production in New Zealand*. Massey University Press, Auckland, New Zealand, 2017
- StatsNZ.** *Primary production – poultry: June 2018 quarter and year – Infoshare tables*. Available at: <https://www.stats.govt.nz/information-releases/primary-production-poultry-june-2018-quarter-and-year-infoshare-tables>. Accessed 26 August 2019. StatsNZ, Tauranga Aotearoa. 2018
- Stringleman H, Scrimgeour F.** *Dairying and dairy products*. Available at: <http://www.TeAra.govt.nz/en/dairying-and-dairy-products>. Accessed 10 February 2019. Te Ara - the Encyclopaedia of New Zealand, 2008
- Tatem AJ, Rogers DJ, Hay SI.** Global transport networks and infectious disease spread. *Adv. Parasitol.* 62, 293–343, 2006
- Vasylyeva TI, Friedman SR, Paraskevis D, Magiorkinis G.** Integrating molecular epidemiology and social network analysis to study infectious diseases: Towards a socio-molecular era for public health. *Infect. Genet. Evol.* 46, 248–255, 2016
- Vermunt J.** Infectious diseases of cattle in New Zealand Part 2 – adult animals. *Surveillance* 27, 3–9, 2000
- Watts JL.** Etiological agents of bovine mastitis. *Vet. Microbiol.* 16, 41–66, 1988
- Whiley H, Van den Akker B, Giglio S, Bentham R.** The role of environmental reservoirs in human Campylobacteriosis. *Int. J. Environ. Res. Public Health* 10, 5886–5907, 2013

- World Health Organisation (WHO).** *The global view of campylobacteriosis: report of an expert consultation.* Available at: https://apps.who.int/iris/bitstream/handle/10665/80751/9789241564601_eng.pdf?sequence=1. Accessed 2 October 2019. World Health Organization, Utrecht, Netherlands, 2012
- Wilson ME.** Travel and the emergence of infectious diseases. *Emerg. Infect. Dis.* 1, 39–46, 1995
- Wintle V, Lepper S.** *Poultry industry.* Available at: <https://teara.govt.nz/en/poultry-industry>. Accessed 10 February 2019. Te Ara - the Encyclopaedia of New Zealand, 2012
- Wulf M, Voss A.** MRSA in livestock animals-an epidemic waiting to happen? *Clin. Microbiol. Infect.* 14, 519–521, 2008

Literature Review

Network Analysis in the Molecular Era: Untangling the Complex Relationships between Host Population Contact Structure, Disease Transmission Dynamics, and Pathogen Evolution

Sabrina S. Greening, Arata Hidano, Ji Zhang, Nigel P. French and M. Carolyn Gates

2.1. Abstract

Understanding how an infectious disease spreads through a population is a fundamental step in developing a successful control strategy, particularly when resources are limited. In recent years, social network analysis has become a popular framework for characterising the complex and heterogeneous contact structures inherent in host populations, with network models providing valuable insights into the transmission dynamics of many important infectious diseases. In the real world, however, multiple transmission pathways amongst multiple host populations make it difficult to capture all the potential contacts that could be contributing to the spread of an infectious disease. This has led to an upsurge in new methodologies that aim at integrating molecular sequence data in network analyses; supporting inferences on the origin, spread, and differentiation of pathogen populations based on the degree of similarity between the genetic sequences of sampled pathogen isolates. This concept, although not new, requires the development of epidemiological methods that can incorporate the evolutionary processes commonly described in population genetics, however despite the increasing availability of molecular sequence data, many molecular epidemiological methods are still in their infancy and have yet to be validated. With this in mind, it is clear that there is need for critical discussions on how pathogen sequence data can be integrated into epidemiological approaches such as social network analysis and network simulation models.

In this review, the focus is on how host population contact networks can be used in complement with pathogen sequence data to (i) make inferences on disease transmission dynamics and (ii) increase our understanding of how population contact structures may be influencing the evolution of pathogen populations, whilst also providing a background into some of the basic phylogenetics concepts that are needed to quantify the genetic relatedness of sampled pathogens. The review begins by summarising the existing literature on network analysis with reference to epidemiological studies that have

successfully used network models to infer disease dynamics in both livestock and human populations, before examining how pathogen sequence data has been used to reconstruct transmission trees and make inferences on the spatial and temporal dynamics of different diseases. Finally, the potential for integrating pathogen sequence data into traditional network analysis is explored, including discussions on the current limitations and future opportunities including consideration for how human behaviour drives the evolution of both host contact structures and pathogen populations; a feature that is often overlooked in traditional infectious disease models.

2.2. Introduction

Over the past few decades, there has been a steady decline in the burden of infectious diseases with Global Health Estimates (GHE) in 2000 attributing 16.4% of deaths to this cause in comparison to only 9.7% of deaths in 2016 (WHO, 2018). Yet despite this trend, the overall number of human infectious disease outbreaks continues to rise (Smith *et al.* 2014) with a growing number thought to be zoonotic in origin (Taylor *et al.* 2001). The recent emergence and spread of zoonotic diseases such as Nipah Virus (NiV), Ebola Haemorrhagic Fever (EHF), Severe Acute Respiratory Syndrome (SARS), and Highly Pathogenic Avian Influenza (HPAI) have exposed gaps in our understanding of how interactions occurring at the human-animal-ecosystem interface are shaping the evolutionary history of pathogens and driving disease emergence and transmission between multiple host species (Malave *et al.* 2010; Kumar and Kumar, 2018; Munster *et al.* 2018).

However, accurately capturing the complexity of these interactions presents several challenges; many of which, call for integrative approaches that have arguably been hindered both by dysfunction in the governance of global health and shortcomings in academic, institutional, and disciplinary silos (Lee and Brumme, 2013). In the meantime, the threat of infectious diseases to both human and animal health continues to rise, accelerated by both an increase in the virulence and resistance patterns seen in many pathogens in response to various anthropogenic and ecological factors that are not fully understood (Lebarbenchon *et al.* 2007; Gottdenker *et al.* 2014; Hendry *et al.* 2017). These challenges have fuelled a growing body of research that aims at integrating molecular sequence data from sampled pathogens and populations contact networks in order to gain understanding of how the structure of host populations and the dynamic interactions between multiple host species influence the emergence and spread of infectious diseases, as well as increasing understanding into the importance of feedback from human interventions on the evolutionary dynamics of pathogen populations.

There are many existing epidemiological tools geared towards understanding how infectious diseases spread through populations; however, in recent years the construction of host contact networks has become an increasingly popular tool with the ability to capture interactions between multiple host species and the different types of contacts that may exist between them (Luke and Harris, 2007; Danon *et al.* 2011; Stattner and Vidot, 2011). Social network analysis (SNA) is the analytical framework often used to describe contact networks which can then be used to build infectious disease models with demographics that more accurately reflect the heterogenous contact structure within the population. Network models, such as these, have provided valuable insights into the transmission dynamics of many important diseases (Keeling, 2005; Bansal *et al.* 2007; Volz *et al.* 2011), helping to guide control and surveillance activities (Eames and Keeling, 2003; Kiss *et al.* 2006a). However, for many diseases multiple contact networks relevant for disease transmission may exist, and it is often difficult to determine which subset of contacts in a real-world outbreak have contributed to the spread of a disease; a process that is often further hindered by both a lack of complete disease data and contact data (Craft, 2015). Therefore, an increasing number of studies are integrating molecular sequence data into traditional epidemiological methods, such as SNA and network models, to provide a greater resolution into the complex transmission pathways of many pathogens (Gardy *et al.* 2011; Inns *et al.* 2017; Campbell *et al.* 2018).

Such molecular epidemiological studies can broadly take one of two standpoints: static or dynamic. In the static approach, sampled individuals who have isolates with a high degree of genetic similarity are considered as clusters, with analyses focusing on determining if any common epidemiological factors exist within a cluster such as the presence of a direct link within the population contact network (Booth *et al.* 2013; VanderWaal *et al.* 2014), or individual-level characteristics including age, sex and specific risk behaviours (Mullner *et al.* 2010; Jaros *et al.* 2013). In contrast, the dynamic approach quantifies the epidemiological relatedness between individuals from whom isolates were sampled by

modelling the genetic mutations; an application that often involves the construction of transmission trees based on genetic sequences of samples. Until recently, many molecular epidemiological studies have taken a static approach, similar to that in many phylogenetics studies, and have been able to infer disease dynamics by determining the tree topology that describes the evolutionary relationship between sampled pathogens (*i.e.*, a phylogenetic tree). For example, Kouyos and colleagues (2010) were able to identify important demographic factors of individuals from a phylogenetic tree containing 11,400 human immunodeficiency virus (HIV) sequences by first determining transmission clusters in the tree topology (Kouyos *et al.* 2010). However, with the advent of whole-genome sequencing (WGS) technology, which can provide a greater resolution in comparison to traditional genotyping methods (Gardy *et al.* 2011; Price *et al.* 2014; Ahlstrom *et al.* 2015) and the emergence of freely-available software such as BEAST (Drummond and Rambaut, 2007), Outbreak Tools (Jombart *et al.* 2014a) FastML (Ashkenazy *et al.* 2012), and MrBayes (Huelsenbeck and Ronquist, 2001), an increasing number of studies are transitioning onto dynamic approaches that have thus far been more common in population genetics as a means of identifying factors associated with genetic divergence (Kühnert *et al.* 2011). While both standpoints have offered promising insights into infectious disease epidemiology, these methods are still in their infancy with many genealogy-based approaches yet to be able to incorporate the complex demographic structures that are known to be an important for infectious disease dynamics. This crossroad presents a timely opportunity to review the developments in analytical methods integrating host contact structures and molecular sequence data and discuss the underlying concepts and assumptions behind each method.

In this review, the current uses of both network data and molecular sequence data in epidemiological research are summarised, with further discussion into some of the underlying principles inherent to many of the methodologies. Before moving on to examine how the two data sources have been integrated; using examples of both human

and animal diseases to highlight some of the limitations that need to be assessed in order to ensure the continued development of robust and reliable approaches.

2.3. Population structure, contact networks, and infectious disease dynamics

The use of infectious disease models in epidemiology has a rich history (Lessler and Cummings, 2016; Brauer, 2017) with Ronald Ross (1857-1932) capturing the fundamental difference between the study of infectious diseases in populations in comparison to other health conditions with the term “dependent happenings”, emphasising the importance of the interactions between individuals with differing disease statuses and disease dynamics (Ross, 1916). Despite this, many earlier infectious disease models were constructed under the simplifying assumption of homogeneous mixing; that is, all individuals in a population mix both uniformly and randomly with each other (Bansal *et al.* 2007). However, with the development of numerous methods capable of incorporating different contacts rates into epidemiological models, it has become increasingly clear just how much heterogeneity in host contact patterns can profoundly shape population-level disease dynamics. For example, one key parameter often used to estimate the transmission potential of a pathogen is the basic reproduction number (R_0): defined as the expected number of secondary infections that are a direct result of a single infected individual interacting within an entirely susceptible host population (Anderson and May, 1992). This parameter has been widely used in epidemic models to predict the speed at which a pathogen may spread through a population and determine the potential size of an outbreak, with pathogen success dependent on having an R_0 that is greater than, or equal to, one (Ridenhour *et al.* 2014; Delamater *et al.* 2019). However, there is often considerable variation in the estimated value for R_0 as a result of different assumptions about the underlying host contact structure.

The impact of the homogeneous assumption was evident in many of the disease transmissions models that were used to estimate important transmission parameters in

the recent SARS epidemic. In the start of the outbreak, original estimates for R_0 were based largely on transmission data from local settings in Guangdong, China, where the outbreak was believed to have been started, with values varying from 2.2 to 3.6 (Meyers *et al.* 2005). With this range of R_0 values, it was predicted that without intervention there could have been anywhere from 30,000 to 10 million SARS cases in the first 120 days (WHO, 2003); however, only 782 cases were actually reported in China during this time period. Large discrepancies in the value for R_0 for SARS were also reported between countries and between different social settings within a country (Xu *et al.* 2004; Meyers *et al.* 2005). Further investigation suggested that these differences were caused from studies estimating R_0 using transmission data from settings such as hospitals or crowded apartment buildings where the contacts between individuals are much closer and more frequent than in the general population, especially considering changes in human behaviour that were likely to have occurred during the epidemic, overall leading to higher transmission estimates (Riley *et al.* 2003; Yu *et al.* 2004). This is just one example that emphasises the influence of host population contact structure on disease dynamics, resulting in a distribution of R_0 values for a single pathogen.

2.3.1. Network theory

One popular method that can be used to describe features within a population contact structure is SNA, with a growing number of studies using network data in disease transmission models to gain an understanding of the influence of contact patterns on transmission dynamics (Christley *et al.* 2005; Keeling, 2005; Shirley and Rushton, 2005a; Parham and Ferguson, 2006). Principally SNA uses graph theory to provide a conceptual framework that can be used to help gather, visualise, and analyse population contact data (Otto and Rousseau, 2002). Basic contact networks are generated with a set of elements, often referred to as nodes, vertices or actors, which represent the unit of interest, and edges or contacts to show the links between them (Figure 2.1). For example, in infectious disease studies nodes often represent an individual (*i.e.*, a human or an animal) or a larger

unit, such as a hospital or a farm, that are connected via different contact pathways that are known to be important for disease transmission such as the transfer of a patient between two hospital or the movement of an animal between two farms. Once the nodes and edges have been defined, networks can be represented and recorded in three ways; (i) simple network diagrams, (ii) mathematical notations that list the network in the form $G = \{n_i, n_j\}$, where G refers to a network made up of sets of interacting nodes (n_i, n_j), and (iii) an adjacency matrix showing $N \times N$ nodes and the number of contacts between. Further information on node attributes may also be collected to determine important risk factors for disease (Friedman and Aral, 2001; Christley *et al.* 2005; El-Sayed *et al.* 2012).

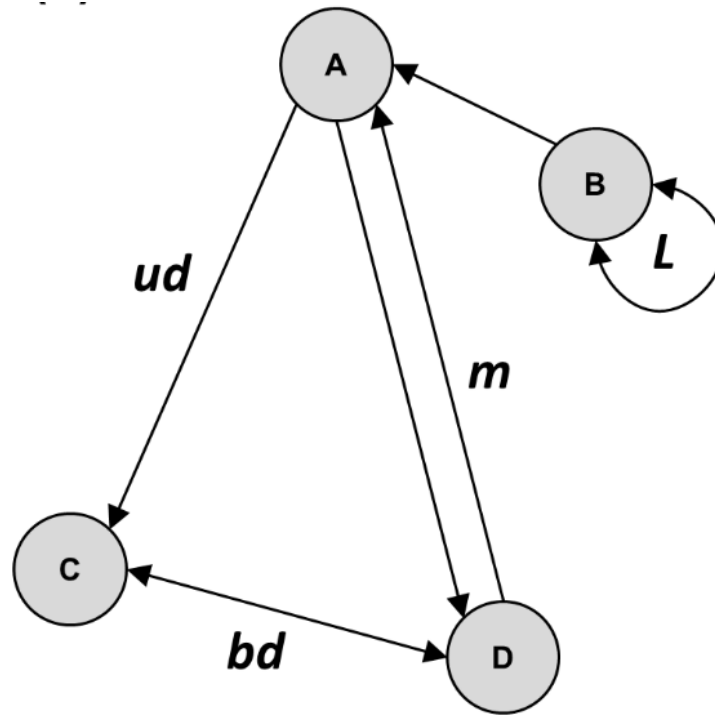


Figure 2.1. (left) Directed network graph $G = \{V, E\}$ showing a set of vertices otherwise known as nodes or actors; A, B, C and, D connected by multiple edges each representing a different relationship including a unidirectional edge (ud), multiple edges (m), bidirectional edges (bd) and loops (L).

SNA and graph theory have been used in many research disciplines including sociology, economics and marketing, psychology, anthropology, biochemistry, neurology, physics, and computer science. Within epidemiological research, it has been used as a tool to help

explore the transmission dynamics of numerous human infectious diseases including; human immunodeficiency virus (Curtis *et al.* 1995; Morris *et al.* 1995; Service and Blower, 1995; Barbosa *et al.* 2000; Bell *et al.* 2002), tuberculosis (Klov Dahl *et al.* 2001; McElroy *et al.* 2003; Cook *et al.* 2007; Gardy *et al.* 2011), hepatitis C (Aitken *et al.* 2004; Brewer *et al.* 2006; Pilon *et al.* 2011), chlamydia (Stoner *et al.* 2000; Wylie and Jolly, 2001), gonorrhoea (De *et al.* 2004) and syphilis (Rothenberg *et al.* 2000; Choi *et al.* 2007; D'Angelo-Scott *et al.* 2015). Many of these early studies focussed on the spread of sexually transmitted infections (STIs) because, unlike many infectious diseases, STIs have a very clear transmission route making contacts easier to trace with a process called 'snowball sampling' whereby an individual in the network recalls all their sexual partners over a given period, these partners are then traced and asked for their partners, and the process is repeated (O'Malley and Marsden, 2008). However, this approach is not without its limitations, often suffering from potential biases if participants cannot recall perfectly every contact made or do not wish to fully disclose all information due to the sensitive nature of a contact.

Other limitations in capturing a full network have been widely recognised with many SNAs constrained by the use of incomplete network data containing both missing nodes and edges (Farine and Whitehead, 2015). For this reason, a variety of methods have been developed to infer synthetic networks (Kiss *et al.* 2006b; Gates *et al.* 2015; Liu *et al.* 2018), which can either be entirely theoretical, with set structural properties, or be created using probability distributions and known attributes from the biological network. However, it is often hard to validate these inferred synthetic networks, and it is often more important to understand why network data may be incomplete, what impact this will have on the network structure, and what the importance is of missing data in terms of the biological question under investigation. For example, missing data are common in networks where the disease dynamics are not fully understood and the host population remains undefined, as is the case with many zoonotic diseases for which there may be multiple

species acting as reservoir hosts with many transmission pathways still unclear (Craft, 2015), making it very difficult to fully capture the properties of the real network. Even in well-defined networks, contact patterns may also change dynamically over time and therefore cannot be adequately represented using traditional static network constructions (Bansal *et al.* 2010; Masuda and Holme, 2013; Pastor-Satorras *et al.* 2015; Enright and Kao, 2018). Given this uncertainty, many approaches can be used to try and assess the precision of estimates in an observed network including bootstrapping (Lusseau *et al.* 2008) and hierarchical models (Cross *et al.* 2012), and despite missing data many networks have been used to great success in infectious disease research.

2.3.2. SNA and infectious disease dynamics

In SNA, a number of networks metrics can be used to characterise individual nodes and the structures between them. For the purpose of this review only measures of ‘centrality’ and ‘cohesion’ will be summarized as more comprehensive reviews of general network concepts are already widely available in the literature (Wasserman and Faust, 1994; Dubé *et al.* 2009). Measures of node centrality were first formalised by Freeman (1978) to help determine the importance of each individual node in a network with measures including node degree, node closeness, and node betweenness (Table 2.1). These simple network characteristics are easy to compute and have been used to characterise a variety of network types, including random graphs, lattice, small-world and scale-free networks (Martínez-López *et al.* 2009). Measures of cohesion focus less on individual nodes and instead are used to determine the level of connectivity over the entire network, allowing the network resilience to be assessed by removing individual nodes or edges. These measures include network density, average path length and the clustering coefficient all of which, offer a comparative way to summarise networks graphs and can be manipulated in network simulation models to understand the impact of topological structures on disease transmission dynamics (Keeling and Eames, 2005; Shirley and Rushton, 2005b; Gates and Woolhouse, 2015).

Table 2.1. Network analysis glossary of terms for measures of centrality and cohesion

Network measure		Definition
Measures of centrality	<i>Degree distribution</i>	Degree distribution is the sum of nodes connected to an individual in the network and can be used to indicate their overall involvement. In a directed graph, it can be further categorised into out-degree (number of contacts originating from a node) and in-degree (number of contacts received by a node).
	<i>Closeness</i>	Closeness is the sum of the length of the shortest paths between an individual and all other nodes in the network.
	<i>Betweenness</i>	Betweenness is the frequency of which an individual is on the shortest path between any pair of nodes.
Measures of cohesion	<i>Network density</i>	Network density is the proportion of all possible contacts an individual could have compared with those that are actually observed in the network.
	<i>Average path-length</i>	Average path-length is the average number of contacts along the shortest path between all pairs of nodes.
	<i>Clustering coefficient</i>	Clustering coefficient is the ratio between the number of connections linking the neighbours of a node and the maximum number of connections that could possibly exist between the neighbours of the node.

Broadly, network simulation models can be divided into two categories: data-driven simulations and simulated networks. Data-driven simulations use construct models with matching properties. The impact of the network characteristics in the observed network, such as the degree distribution, to structure on the spread of disease within the population can then be determined, often using a process called network re-wiring to identify which features present in a biological network have the biggest impact on disease spread, as well as being able to assess the importance of an individual in the network by systematically

removing a node with given properties (Green *et al.* 2006; Mossong *et al.* 2008; Read *et al.* 2008; Danon *et al.* 2011). In comparison, simulated networks stochastically simulate entire networks based on a minimum set of rules which can either then be compared to the observed network, a random network, or be used in isolation to study the influence of different network features. For example, a study by Ames and colleagues (2011) used stochastic simulations to generate artificial networks with different degree distributions, clustering coefficients, and average path lengths in order to investigate the influence of these network properties on the spread of disease. Study results showed that these three attributes contributed to over 98% of the variation in endemic disease levels and can be used as robust measures to predict the spread of disease through a network (Ames *et al.* 2011). Simulated networks have also been used to characterise different types of networks based on the properties they share and their impact on disease spread such as small-world networks that are identified by a high clustering coefficient and short average path lengths; two attributes that are thought to increase the speed of transmission due to a relatively small number of ‘unclustered’ or random links, and cause a rapid depletion of local susceptible contacts (Watts and Strogatz, 1998; Holme and Kim, 2002; Newman, 2003; Volz *et al.* 2011). In comparison, scale-free networks are characterised by a small number of individuals that have a disproportionately large number of contacts in the network (*i.e.*, they have a highly skewed degree distribution) which results in elevated values for R_0 even when the transmission probability is low, although it is important to note that not all highly skewed distributions are scale free (Barabási and Albert, 1999; Newman, 2002; Woolhouse *et al.* 2005; Kiss *et al.* 2008).

For many populations, a simulated network is necessary due to the limitations in observing real-world contacts. However, an increasing number of databases, utilising a range of technologies, are being constructed to capture population dynamics, particularly in livestock systems where the fostering of regulations on the reporting of livestock movements has resulted in an influx of available data making it relatively easy to quantify

the frequency of direct animal movements onto and off farms. Much of the research using livestock movement data followed the 2001 foot-and-mouth disease (FMD) epidemic in the United Kingdom (UK) as many of the early models had a limited predictive ability due to the assumptions around the contact structure of farms (Woolhouse and Donaldson, 2001). Subsequently, livestock movement networks for both sheep and cattle in the UK have been described and analysed using SNA to try and identify network structures and risk factors that were important in the FMD outbreak. This has helped to identify potential targets for control including key farms, auction markets, dealers and slaughterhouses which showed high connectivity and would have been important in the initial spread of the disease (Ortiz-Pelaez *et al.* 2006; Robinson and Christley, 2007).

In addition to FMD, SNA has been used to study other livestock diseases such as bovine tuberculosis (Corner *et al.* 2003; Gilbert *et al.* 2005; Carrique-Mas *et al.* 2008; Christley *et al.* 2011; Gates *et al.* 2013; Grear *et al.* 2014), equine influenza (Christley and French., 2003), scrapie (Kao *et al.* 2007), avian influenza (Dent *et al.* 2008; van Kerkhove *et al.* 2009; Martin *et al.* 2011; Fournié *et al.* 2013; Poolkhet *et al.* 2013), brucellosis (Savini *et al.* 2017) and bovine viral diarrhoea virus disease (Tinsley *et al.* 2012). However, despite increasing access to network data, many of these studies suffer from the same limitations resulting from knowledge gaps or mismatches between relevant contact patterns and disease dynamics (Craft, 2015). To overcome some of these challenges many approaches have been taken, including both observational and experimental studies, however, more recently an increasing number of studies are using molecular sequence data to reconstruct pathogen phylogenies and infer transmission dynamics (Grenfell *et al.* 2004; Colijn and Gardy, 2014; Kamath *et al.* 2016).

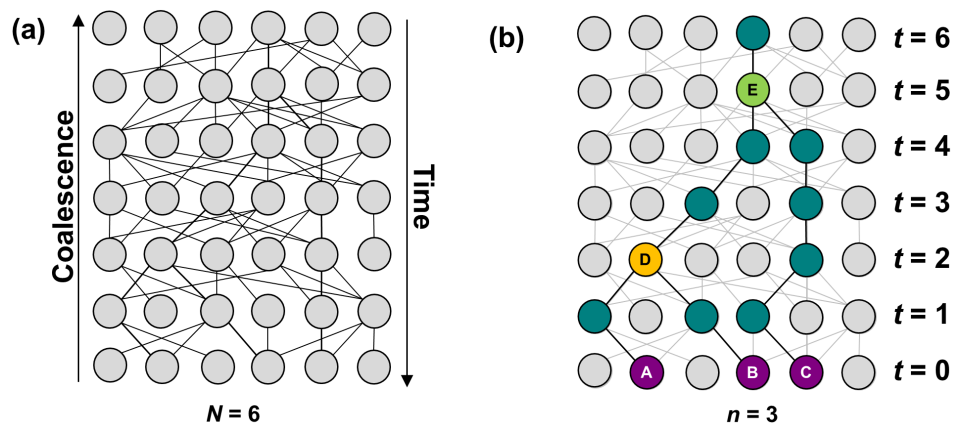
2.4. Pathogen population dynamics and infectious diseases

Over the last decade, molecular epidemiology is a discipline that has rapidly expanded (Eybpoosh *et al.* 2017), with many earlier studies using molecular typing methods as a

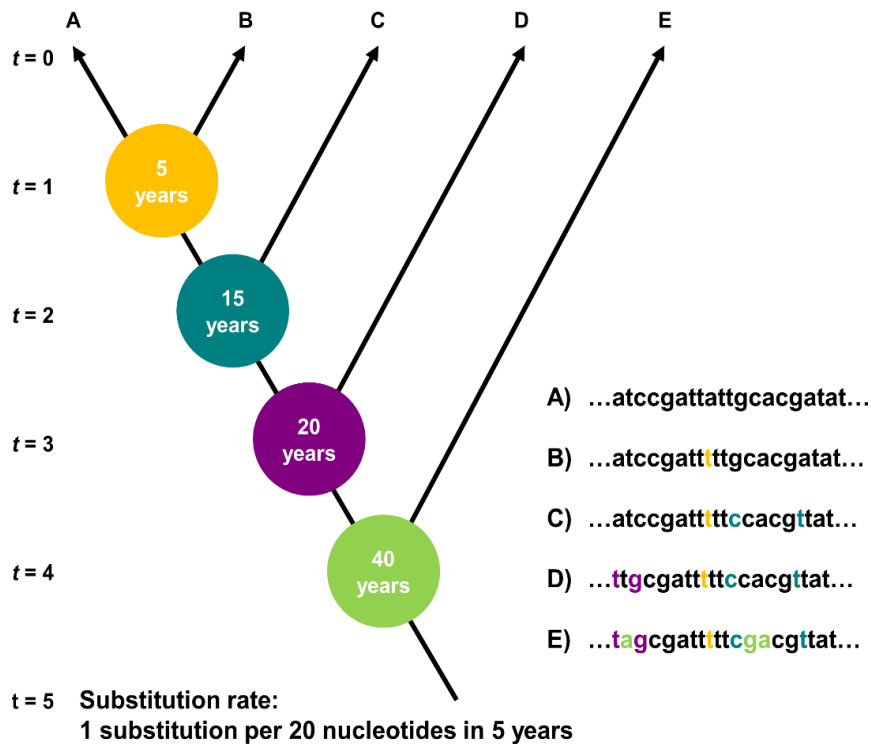
diagnostic tool with no consideration for the evolutionary information contained within the sequence. Used in this way, genotyping data can provide a more detailed case definition as strain types within the same pathogen can be identified and more meaningful inferences on disease epidemiology can be made with different strain types often having distinct transmission patterns. More recently with developments in population genetics, computational biology and evolutionary modelling, many molecular epidemiological studies can use typing data to investigate the distribution, dynamics, and determinants of health and disease in populations (Field *et al.* 2014). For the purpose of the review no comparison has been made between the different typing methods, with comprehensive reviews already provided in the literature (Ranjbar *et al.* 2014), however it is important to note that in epidemiological studies the selection of specific molecular marker among various typing methods depends on the purpose of the analysis, and the spatial and temporal extent to which the study aims to make inferences (van Belkum *et al.* 2001). Nevertheless, the basic principles used remain the same regardless of the typing methods; that is, that the amount of differences between pathogen sequences represent the temporal divergence from a common ancestor (Restif, 2009; Field *et al.* 2014). This principle is rooted in coalescent theory (Kingman, 2000) and tree-thinking approaches (Box 1) that often require the incorporation of evolutionary information such as the mechanism of change in the genetic marker, and the rate of evolution. Acquiring this information is often the limiting factor in many molecular epidemiological studies due to the relatively coarse molecular data provided by some typing methods (Riley, 2004). However, with the advent of high-throughput sequencing technologies providing unprecedentedly high-resolution WGS data, many studies have been able to disentangle complex epidemiological phenomenon that could not be previously identified using conventional methods.

The differences in the resolution of molecular data has led to two broad standpoints in molecular epidemiological research with many earlier studies that used more traditional

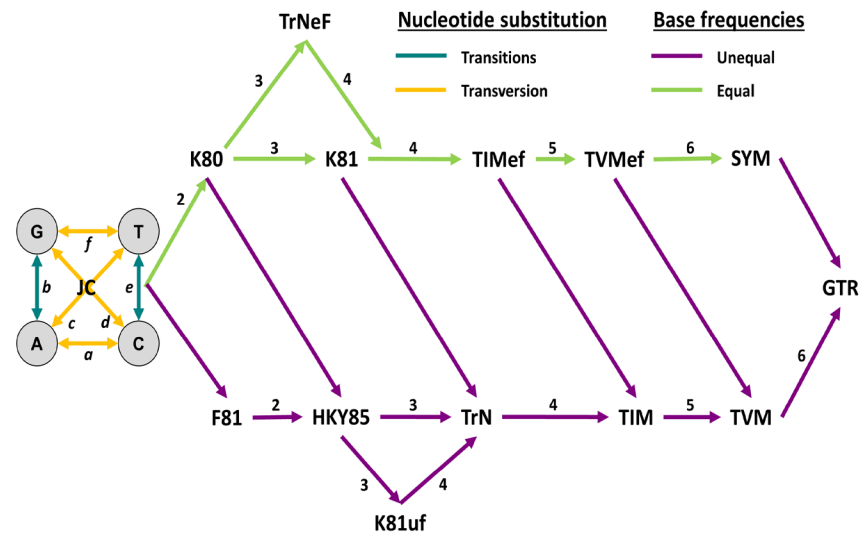
typing methods treating data in a static manner, with the exception of a few applications such as stepwise mutation models (Aandahl *et al.* 2012). Many static approaches have a history in phylogenetics and focus on identifying the topology of evolutionary relationships so that sampled pathogen isolates can be categorised into different groups based on their genetic closeness (Pagel, 1999). In comparison, WGS data is often treated in a dynamic manner, taking approaches from different disciplines such as population genetics to make inferences on the dynamics of pathogen populations at different spatiotemporal scales. Although it is acknowledged that many of the underlying concepts and assumptions in phylogenetics are also prerequisite to the dynamic approaches used in population genetics such as the molecular clock hypothesis (Box 2) and substitution models (Box 3).

Box 1 – Coalescent theory

The coalescent model was first formulated by Kingman (1939-present) as a mathematical model to randomly construct genealogies backwards in time (Kingman, 2000). The figure above shows a schematic of the most basic coalescent, with (a) representing the complete genealogy of a population where $N = 6$, (b) represents a sample within the population ($n = 3$; A, B, and C) that has been traced back along their descendent line until their most recent common ancestor (MRCA). In a traditional coalescent model, a pair of individuals are randomly selected and merged to form a single ancestor from the present (generation $t = 0$) to the past (generation $t = 1$ to $t = 6$). The probability of two individuals randomly selected from the population sharing a common ancestor at generation $t = 1$ is $1/N$ and the probability of not finding a common ancestor is $1 - 1/N$ (Aldous *et al.* 2001). This can be easily extended to describe the probability that any two randomly selected individuals from the population sharing a MRCA in a past generation. Many extensions have been made to relax assumptions made in the basic coalescent model; however, the concept that a distribution of generation intervals and hence the time to MRCA can be described as a function of population size is fundamental to any of these models.

Box 2 – Molecular clock hypothesis

The molecular clock hypothesis was originally introduced in the 1960's by Zuckerkandl and Pauling who proposed that genetic mutations accumulate at constant rate over time (Zuckerkandl and Pauling, 1962). This assumption allows for the estimation of historical dates of events such as the time to the most recent common ancestor (TMRCA), as shown in the figure above, meaning in an epidemiological investigation, the evolutionary history of pathogens can be aligned with epidemiological events that have occurred in comparable time. Molecular clocks can be broadly divided into strict molecular clocks and relaxed molecular clocks, with the former assuming that the evolutionary rate is constant over every branch of a tree whilst relaxed molecular clocks allow the evolutionary rate to vary across tree branches in different ways depending on the model. For example, in an uncorrelated relaxed molecular clock model the evolutionary rate follows a specific underlying distribution such as exponential, log-normal or inverse Gaussian distribution (Drummond *et al.* 2006). To assist in molecular clock model selection, standard statistical methods such as likelihood ratio test and Akaike information criterion (AIC) can be used (Arbogast *et al.* 2002); however, in many cases it has been found that the basic assumptions underlying the chosen molecular clock have been violated (Kumar, 2005), with a large number of studies resorting to unrooted trees instead.

Box 3 – Substitution Models and Tree Reconstruction

The topology of phylogenetic trees is constructed either based on the genetic distance between samples or modelling the nucleotide patterns in each site. The simplest genetic distance is often the p -distance. This metric is, however, often inappropriate when the time span between sequence sampling is long and the substitution rate is high. This is because a single nucleotide site can experience multiple substitution events which might be masked by more recent substitutions and cannot be accounted for by this simple distance measure. Such underestimation can be adjusted by either explicitly formulating a relationship between the observed and expected distance or constructing a likelihood function that describes the probability of observing a set of nucleotides at each site of alignments given the expected distance. Both methods are based on different substitution models that specify the probability of substitutions between each of the four nucleotides (*i.e.*, transitions versus transversions) and the base composition, as summarised in the figure above (a, b, c, d, e, f).

Different substitution models have different parameters to be estimated. The figure above shows how different substitution models are nested within each other including the Jukes and Cantor model (JC) (Jukes and Cantor 1969), the Felsenstein model (F81) (Felsenstein 1981), the Kimura model (K80/1) (Kimura 1980), the Hasegawa-Kishino-Yano model (F84) (Hasegawa *et al.* 1985), the Tamura-Nei model (TrN) (Tamura and Nei, 1993), the transitional model (TIM), the transversion model (TVM), the symmetrical model (SYM) (Zharkikh 1994) and, the general time reversible model (GTR) (Rodríguez *et al.* 1990) with either equal base frequencies (ef), highlighted by the green arrows, or unequal base frequencies (uf) as highlighted by the purple arrows, and the number of substitution classes in each model shown by the corresponding number on each arrow.

2.4.1. Pathogen phylogenetics

Phylogenetics studies aim to reconstruct the evolutionary history or phylogeny of related organisms based on their evolutionary similarities and differences. For the purpose of this review, many of the detailed steps that preclude a phylogenetic analysis, such as multiple sequence alignment, are not discussed as more comprehensive reviews already exist in the literature; although, it is worth recognising the importance of each step and the impact they have on phylogenetic reconstruction (Harrison and Langdale, 2006; Yang and Rannala, 2012). Nevertheless, the main focus of many molecular epidemiological studies is in describing the inferred phylogeny, often in the shape of a phylogenetic tree, and investigating its association with disease dynamics. For example, the genetic similarity between isolated pathogen sequences can be assessed with clustering algorithms, using both hierarchical or margining approaches, to identify the origin of epidemic strains by comparing the sequences of epidemic and non-epidemic isolates (Kühnert *et al.* 2011).

There are various methods that can be used to construct a phylogenetic tree but mostly, they can be characterised into two broad classes based on either discrete character or distance-based algorithms. Discrete character methods such as maximum parsimony, maximum likelihood (Felsenstein, 1981) and Bayesian approaches (Rannala and Yang, 1996) generally search for an optimal tree topology that fits the observed data based on some criteria (*e.g.*, minimum substitution or likelihood), whilst distance based methods, such as neighbour-joining and Unweighted Pair-Group Method with Arithmetic mean (UPGMA), use algorithms to construct a tree by sequentially connecting nodes based on the fraction of sites that differ between two sequences (Rizzo and Rouchka, 2007). There are many different advantages and disadvantages to each of these methods (Whelan *et al.* 2001; Holder and Lewis, 2003; Steel, 2005; Simmons, 2014) and how to select the most appropriate method is highly debated with a number of considerations to account for such as the complexity of the evolutionary model being used and computation time (Douady *et al.* 2003; Bos and Posada, 2004). However, regardless of the method used, it

is important to bear in mind that the homologous sites of genes or amino acids should be compared to provide a meaningful tree that can be used in epidemiological investigations (Lemey *et al.* 2009). Further uncertainty in tree reconstruction also makes it impossible to guarantee that the identified tree topology reflects the true phylogeny and although standard statistical methodologies such as bootstrapping and jack-knifing are available to evaluate the reliability of a constructed tree, they typically assess the reproducibility of branches and not the accuracy of the topology. This means that some care must be taken when interpreting trees in an epidemiological context as many different scenarios can reproduce the same tree topology whilst many different topologies can also equally explain the observed data.

Despite the uncertainties in the topology of a phylogenetic trees, they have proven to be a valuable tool in determining the spatiotemporal spread of many pathogens (Osmani *et al.* 2014; Rosendal *et al.* 2014). For example, the global spread of HPAI has sparked numerous investigations into the role of wild bird migration and poultry trade in the emergence and circulation of the virus with phylogenetic analyses being used to investigate dynamics on a range of geographical scales (Liang *et al.* 2010; Lewis *et al.* 2015; Tian *et al.* 2015; Briand *et al.* 2017). Other studies have adapted a range of methodologies rooted in multiple disciplines to incorporate phylogenetic data including multidimensional scaling (Bergholz *et al.* 2010; Carrel *et al.* 2012; Wehner *et al.* 2014; Carrel *et al.* 2015), spatial Eigenfunction analysis (Diniz-Filho *et al.* 2013; Tedersoo *et al.* 2013), spatial autocorrelation (Pybus *et al.* 2012; Garbelotto *et al.* 2013; Omedo *et al.* 2017), analysis of similarities (Voss *et al.* 2007), permutational analysis of variance (Grange *et al.* 2015), multivariable linear-mixed models (Zhang and He, 2013; Tamminen *et al.* 2019), generalized additive model (French *et al.* 2005), and quadratic assignment procedures (Marquetoux *et al.* 2016). However, many of these approaches are limited in their ability to incorporate evolutionary processes and while the topology of phylogenetic trees provides useful information about disease epidemiology, the traditional

phylogenetic approach cannot be used to estimate quantitative parameters that can be readily used in epidemiological models. In contrast, many genealogy-based approaches, looking at genetic diversity and its change over time, have the ability to produce quantitative estimates on the demographic history of pathogen populations in the past and therefore have been used to infer infectious disease dynamics (Welch *et al.* 2005; Kühnert *et al.* 2011).

2.4.2. Genealogy-based modelling

At the centre of genealogy-based approaches are different methods that can estimate population divergence time or the time to the most recent common ancestor (TMRCA). These methods have vastly different frameworks which require making different assumptions about the presence of evolutionary processes, such as recombination and demographic structure; however, the TMRCA can provide insight into different disease dynamics. For example, if the TMRCA is homogeneous and recent back in time, then this may suggest the recent introduction of their ancestor or the presence of a transmission bottleneck event such as an outbreak in which only a small portion of infected individuals were responsible for the spread of a disease (Romano *et al.* 2010; Bataille *et al.* 2011; Famulare and Hu, 2015; Kamath *et al.* 2016). The estimated TMRCA of pathogen populations may also provide further epidemiological inference such as the plausibility and direction of transmissions. For instance, if the TMRCA between two sampled isolates is prior to the age of both individuals from which the isolates were sampled, then transmission could not have occurred between two as neither of the individuals were born at the estimated TMRCA and hence could never have harboured a common strain (Didelot *et al.* 2012).

The increasing use of genealogy-based approaches has been facilitated by the development of Bayesian approaches that can readily incorporate evolutionary processes into an existing phylogenetic framework by specifying genealogy-based models as prior

information for tree topology, the implementation of which is easily accessible in freely available computer software. However, while genealogy-based models have provided valuable information on the dynamics of many infectious disease, the assumptions underlying many of these approaches are easily violated, and how such assumptions and sampling bias affect the estimates is not yet fully understood. Moreover, current genealogy-based models are not able to consider the complex population demographics which are important components in epidemiological-based models. For these reasons, there has been a strong interest among epidemiologists to develop a unified method that can incorporate detailed population contact structures as well as high-resolution pathogen sequence data.

2.5. Epidemiological models incorporating genetic data

One of the first pivotal papers combining epidemiological data into a phylogenetic analysis was that by Cottam and colleagues (2008) that reconstructed the transmission pathways of FMD by using known contacts to substantially reduce the number of possible transmission trees that were consistent with the genetic data (Cottam *et al.* 2008). Transmission tree reconstruction begins in a similar process to constructing a phylogenetic tree whereby sampled pathogen isolates are genotyped, using molecular markers to identify genetic variants, and the genetic distance between isolates calculated to quantify the relatedness between each sample (Whelan *et al.* 2001; Edwards, 2009; Yang and Rannala, 2012). Several parameters and their estimators exist to measure the genetic distance between samples with some metrics being more appropriate than others when taking into consideration the underlying geometric and evolutionary assumptions inherent to phylogenetics with many detailed reviews discussing the appropriateness of different methods (Jombart *et al.* 2011; Joly *et al.* 2015). The observed genetic distance can then be matched to the epidemiological relationships between infected hosts with direct transmission ruled out by using either a fixed threshold for substitutions or by using evolutionary models to estimate the expected genetic distance between isolates

(Gonzalez-Candelas *et al.* 2003; Jombart *et al.* 2014; Croucher and Didelot, 2015; Lau *et al.* 2015). Simply put, transmission trees can be inferred based on the assumption that the genetic relatedness between pathogen isolates will be shorter if the hosts from which the isolates were sampled from were closer in the chain of transmission (Ypma *et al.* 2012; De Maio *et al.* 2016). These techniques based on pathogen phylogenies have been used in numerous infectious disease outbreaks with increasing popularity for foodborne pathogens for which it is often vital to be able to trace the origin of an outbreak across different geographical and temporal scales (Deng *et al.* 2016; Dunn, 2016; Nadon *et al.* 2017; Pightling *et al.* 2018; Brown *et al.* 2019).

In addition to combining data, several studies have also used epidemiological and genetic models in parallel in order to compare the results. For example, Scarpino and colleagues (2015) applied two Bayesian frameworks to the Ebola outbreak in 2014 using genetic data to reconstruct transmission taking into account the genetic variation, collection dates, duration of pathogen colonization and time interval between cases. Results from both models were found to be consistent to a network-based compartment model that accounted for the clustering of contacts between individuals (Scarpino *et al.* 2015). However, despite the success of these approaches, it is increasingly clear that the genetic relatedness between pathogens cannot always be explained by a clear epidemiological link, and with many of these methods relying on densely sampled host populations and a rapidly mutating pathogen that allows for discrimination between pathogen genomes on an epidemiological timescale (Campbell *et al.* 2018), the impact of sampling bias on these estimates has not been fully understood (Hidano and Gates, 2019). Furthermore, only a limited number attempt to distinguish between the mode of transmission and the transmission route between infected hosts; a differentiation important for effectively controlling the spread of disease. For this reason, many studies are moving towards approaches centred around epidemiological modelling in order to achieve a greater

resolution into disease dynamics, with a growing number of studies using SNA frameworks to integrate host contact networks into a phylogenetic analysis.

One of the seminal papers incorporating heterogeneous host contact patterns into simulation models was that by Lemey and colleagues (2014). This study used a phylogeographic approach to compare the observed peaks for influenza H1N1 in 14 air communities to those derived from compartmental susceptible-infectious-recovered (SIR) simulation models with different transmission parameters (Lemey *et al.* 2014). Phylogeography is one field taken from phylodynamics, a framework that unifies evolutionary biology, epidemiology, and immunodynamics (Grenfell *et al.* 2004), which mainly studies spatiotemporal disease dynamics based on genetic information (Holmes and Grenfell, 2009). Using this framework, the study by Lemey and colleagues (2014) highlighted that the use of transmission coefficients derived from solely phylogeographic models did not produce results that matched to the observed peak, however models that incorporated parameters derived by using both phylogeographic and epidemiological models showed a high correlation to the observed data, revealing that air transportation and the population size of the transport destination were important risk factors for transmission.

Further use of detailed contact information has provided insights into the contribution of different contacts towards the spread on disease. For example, a study by Young and colleagues (2017) evaluated the spatial correlation of avian influenza isolates with various models of geographical distance including Euclidean, road network, road network via intervening live bird markets and least-cost paths. Seventy-three sequenced avian influenza isolates were used in the phylogenetic analysis with isolates from backyard poultry found to be associated with least-cost path distance, whilst isolates from commercial farms were more strongly associated with road network distance; suggesting

that the role of wild birds and human-mediated transmission differ between the different systems of poultry production (Young *et al.* 2017). In a more recent study by Tamminen and colleagues (2019), information from environmental sampling of a highly virulent strain of verotoxigenic *Escherichia coli* (VTEC) from 80 farms was combined with information from farmer questionnaires and whole genome sequencing to investigate different risk factors; providing insight into the importance of ongoing local transmission mechanisms for the spread of VTEC and disease persistence in the population (Tamminen *et al.* 2019). Similar studies can be found throughout the literature that use sequenced pathogen data to identify important risk pathways (French *et al.* 2005; Booth *et al.* 2013; Jaros *et al.* 2013; Alkhamis *et al.* 2017); with many focusing on livestock networks on account of the development of many databases aimed at capturing livestock movements as well as different technological advances such as remote sensors that allow researchers to capture different types of contacts (Handcock *et al.* 2009). The successful integration of these two complementary data sources, alongside advances in network modelling, has also led to a growing number of studies looking at the influence of host contact structures on the evolutionary potential and trajectories of pathogen populations in order to comprehend the risk of disease and help design effective prevention and control strategies (McDonald and Linde, 2002; Barrett *et al.* 2008; Robinson *et al.* 2013; Leventhal *et al.* 2015).

2.5.1. The coevolution of host contact structures and pathogen traits

It has long been known that the spatial structure of pathogen populations can influence the selection pressure on specific pathogen traits (Ewald, 1993) and the coevolution between host and pathogens (Thompson, 1999) however, the extent to which network structures drive the selection for pathogen traits remains largely unclear. Numerous models such as the geographic mosaic theory of coevolution aim to capture the effect of spatially structured populations on the co-evolutionary dynamics of pathogens (Lively, 1999; Thompson and Cunningham, 2002). In a study by Buckee and colleagues (2004)

stochastic spatially heterogeneous models were used to investigate the effects of different host contact networks on pathogen strain diversity and dynamics. The study showed that an increasingly diverse pathogen population emerged as contacts between hosts become more localized. Whilst not eluding to any individual characteristic, these results imply that host contact network structure plays a significant role in mediating the emergence of varying pathogen population structure (Buckee *et al.* 2004).

However, the relationship between host population contact structure, evolutionary dynamics and pathogen traits are not always so clear. For example, the trade-off between two pathogen traits; virulence and transmissibility (Keeling, 2000; Kao, 2006; Lin *et al.* 2016), has been shown to be influenced by a number of properties in a host population as a consequence of a pathogen's adaptive response to maximize its R_0 (van Baalen, 2002). Network models have shown that in highly clustered populations, transmission rates will increase even as virulence decreases in order to avoid the rapid depletion of susceptible hosts and pathogen extinction (Read and Keeling, 2003; Boots *et al.* 2004). Whilst in populations with a higher proportion of contacts between clusters there is a reduced risk of depleting susceptible hosts, increasing the selection for virulence and reducing the transmission rate. This relationship shows how host population contact structures may be acting as major selection pressures for adaptive pathogen traits however, this relationship has been debated with other models considering the level of virulence to be coincidental (Levin and Bull, 1994; Weiss, 2002; Ebert and Bull, 2003), or dependent on the within-host competition between pathogen strains (Alizon *et al.* 2009). This uncertainty emphasises the importance of understanding the link between host contact networks and pathogen dynamics in order to effectively capture disease transmission; however, many of the current evolutionary models have a limited ability to account for host contact structure whilst many disease transmission models neglect pathogen evolutionary dynamics (Buckee *et al.* 2004; De Smet and Marchal, 2010; Valente, 2012).

2.6. Current limitations and future challenges

A number of additional challenges also exist that further limit understanding about the link between host population dynamics, epidemic processes, and pathogen evolution. Besides the many constraints in the epidemiological and evolutionary models currently available (Frost *et al.* 2015; Metcalf *et al.* 2015), there is also a general lack of consensus, consistency and validation for many of the methods being developed (Firestone *et al.* 2019); often driven by both a lack of quality and quantity of suitable data to inform the models. However with an ever-increasing mass of sequence data becoming available, many of these limitations are now starting to be addressed in the literature alongside the use of other big data sources that are facilitating the development of models that aim to account for additional factors that may be driving disease transmission such as human behaviour changes and host dynamics.

2.6.1. Missing or imperfect contact network data

Many of the limitations in network-based methods is due to the unobserved individuals in the population leading to both missing nodes and edges. The challenges in collecting network data have been previously discussed (Mikolajczyk and Kretzschmar, 2008; Eames *et al.* 2015) however, despite it being a common problem, it is often difficult to quantify the impact of analysing incomplete network data on network-level statistics and epidemiological inferences (Kossinets *et al.* 2006). One previous study by Pfeiffer and colleagues (2015) has used a simulation algorithm to estimate the required fraction of a network that needed to be sampled to ensure population estimates of a known precision however, the estimate still relied on being able to fully characterise the population at risk; which can be particularly difficult in isolated wildlife populations or for a disease involving multiple host species (Pfeiffer *et al.* 2015). A recent review by Craft (2015) instead highlights important steps that should be considered when trying to capture contact networks in order to ensure that they can be incorporated effectively into infectious disease studies including giving careful consideration to how a contact is

defined, how an infection may alter the contact structure, how an observed network may be scaled-up to a biologically relevant size, and how contacts between multiple host species can be captured (Craft, 2015). Nevertheless, for current network-based approaches to continue to develop, it is clear that improvements in the collection of network data is required, eliminating any uncertainty around inferring missing nodes and edges.

2.6.2. Modelling dynamic networks

With continued improvements in network data, more realistic transmission models that have captured all the relevant contacts that may be playing a role in disease transmission can be developed; helping to improve epidemiological inferences. However, further considerations are also needed to overcome limitations arising from the use of static contact networks whereby nodes and edges remain fixed for the duration of a model (Bansal *et al.* 2010). This assumption disregards any changing dynamics within a population which may have a huge impact on the spread of a disease depending on the timescale over which the pathogen spreads (Volz and Meyers, 2007), and a number of studies have shown the effects of dynamic network properties, such as the regularity and duration of contacts, on disease spread (Fefferman and Ng, 2007; Read *et al.* 2008; Smieszek, 2009). However, these studies rely on good quality longitudinal data sets that provide information about the timing, identity and duration of contacts which is often limited. Developing the necessary processes in a disease transmission model would also require a basic understanding of the underlying mechanisms. For example, to include a feedback loop that considers the influence of a disease outbreak on the population contact pattern, it would be crucial to have some understanding of patterns in human behaviour and how they influence disease spread (Funk *et al.* 2010). Model feedback systems such as these would not only consider ‘how’ and ‘why’ a disease is present in a population but also ‘what next’, with many of the dynamic network models in the literature investigating how changes in human behaviour in response to a disease outbreak alter disease

transmission patterns (Shaw and Schwartz, 2008; Prado *et al.* 2009; Marceau *et al.* 2010; van Segbroeck *et al.* 2010; Jolad *et al.* 2012). Host behaviour has also been shown to play an important role in animal populations with links between host behaviour, host demography, and the transmission of infection thought to be important in determining the impact of disease at different stages of population growth and decline (De Castro and Bolker, 2005; Silk *et al.* 2019). However, capturing both human and animal behaviour, particularly in response to an event such as a disease outbreak, is not easy and many of the concepts and methodologies needed to approach these research questions are not widely familiar to the epidemiology community or those developing the models.

2.6.3. Phylogenetic complexities

The same is true when handling molecular sequence data in disease models as it requires a basic understanding of pathogen evolution and population genetics. This barrier could somewhat explain why pathogen evolution until recently has been largely neglected in epidemiological models, but also highlights the importance of multidisciplinary research teams especially if current methods using pathogen sequence data to make epidemiological inferences are to be extended and include evolutionary complexities such as changing mutation rates, selection, re-assortment, recombination and within-host variation (Kretzschmar *et al.* 2010; De Maio *et al.* 2016; Campbell *et al.* 2018). The incorporation of feedback loops is also very important to solve some of these problems as infectious disease dynamics are further affected by changes in pathogen characteristics such as virulence, infectiousness, and fitness, which may be influenced by disease interventions. Further expertise is also required in all stages of a phylogenetic analysis in order to avoid some of the potential artefacts that can arise during each step, starting from errors in the genome sequencing to genealogy-based models and tree building exercises, all of which can distort inferences made about historical evolutionary events (Stevens and Schofield, 2003). Nevertheless, the increasing accessibility of sequence data alongside the advent of freely-available computer software may result in the misuse of sequence data in

which important assumptions underlying phylogenetic approaches may be violated particularly in cases where the effect of such assumptions and sampling bias on the estimates is not yet fully understood. For example, one key assumption grounded in coalescent theory is that sampled pathogens are all tips on a phylogenetic tree and the internal nodes correspond to a coalescent event, the timing of which depends on the generation interval and with-in host evolutionary dynamics (Yang *et al.* 2014), but none of the samples at the tip are direct ascendants of other samples. This assumption is unlikely in an epidemic and the nodes in a transmission tree correspond to a transmission event meaning that both ancestors and descendants may both be present in an outbreak resulting in incomplete lineage sorting, erroneous trees and, biased transmission estimates (Gavryushkina *et al.* 2014; Didelot *et al.* 2016).

Another simplification that is often made in evolutionary models is the mechanisms by which sequence variation arises. The clonal model underlines much of bacterial population genetics, maintaining that in the absence of sexual processes, chromosomal variation arises through random *de novo* mutations which can then be passed on to following generations by processes of vertical transmission, with distinct lineages arising from the accumulation of single mutations (Jackson *et al.* 2011; Acuna-Hidalgo *et al.* 2016). However, with advances in sequencing technology came a growing recognition for other evolutionary mechanisms including the role of multiple horizontal DNA transfer processes such as conjugation, transduction and natural transformation (Ochman *et al.* 2000; Gogarten and Townsend, 2005; Thomas and Nielsen, 2005; Soucy *et al.* 2015). However, despite the importance of these processes in generating genetic diversity, the direct effect of spatial heterogeneity and contact patterns on these mechanisms is still largely unknown and methods to quantify the correlation between differences in population structures to the differences in pathogen sequences are limited (Metcalf *et al.* 2015). Therefore the focus of many evolutionary models is often the diversity arising from *de novo* mutations which itself can be weighted by a number of uncertainties due to the

unknown influence of different factors on the mutation rate including biological and geographical factors such as the transmission mode, host species and environmental stresses (Messinger and Ostling, 2009; Streicker *et al.* 2012; Maharjan and Ferenci, 2015), time-dependent heterogeneous evolution rates (Biek *et al.* 2015; Ho and Larson, 2006), intra-genome heterogeneity of mutation rates, different disease phases such as latent periods, dormancy and endospores (Ford *et al.* 2011; He *et al.* 2013), and hyper-mutation (Köser *et al.* 2012). Some models have been developed to allow for differences in the mutations rates either over time or sites in genome (Bielejec *et al.* 2014); however, many of the current models are constructed using a Bayesian framework and thus require prior information the impact of which may seriously bias the results particularly in data limited settings (Fourment and Holmes, 2014).

Similarly, a lack of data also makes it difficult to validate many of the current methods integrating pathogen sequence data and epidemiological data. In order to support decision making, methods must be transparent and capable of producing accurate predictions, however, with a lack of real-world examples, many experimental models are yet to be tested (Bansal *et al.* 2007; Craft, 2015; VanderWaal *et al.* 2016). This further highlights not only the importance of data collection but also in building and maintaining strong lines of communication across research disciplines in order to promote transparency and innovation (Strober, 2006), fuel discussion around the challenges of integrating multiple data sources such as how to weight evidence from different datasets and handle the dependencies between them (De Angelis *et al.* 2015), and work towards finding a consensus between all the methodologies currently being developed. Further discussions considering how to critically assess complex models would also aid researchers when deciding on the most appropriate models to use, which can be considered a skill in itself, and is dependent on the research question, the quality of the data, and the desired output (Keeling and Rohani, 2008; Grant *et al.* 2020).

2.7. Conclusion

The application of network-based approaches in infectious disease epidemiology has led to invaluable insights on how heterogeneity in host contact structures affects disease dynamic; however, it is clear that disease dynamics are dependent on more than just the underlying population contact structure, with the co-evolution of pathogens playing a major role in the emergence and spread of a disease. Unified approaches integrating both epidemiological and evolutionary information will undoubtedly provide a greater understanding into both infectious disease dynamics and pathogen evolutionary trends, with current molecular epidemiological approaches having successfully been used to identify important disease risk factors and infer pathogen transmission. The additional use of network-based models has also highlighted how population contact structures can act as constraints on the evolutionary behaviour of pathogen populations. However, the continued development of these models is constrained by a lack of understanding of how host contact structures may be driving pathogen differentiation and trait selection. This knowledge gap emphasises the need to build long-lasting multidisciplinary relationships in order to not only develop methodologies that are robust and transparent whilst also maintaining the assumptions underlying each approach, but also to effectively integrate additional data that will be crucial to understanding infectious disease dynamics.

2.8. Acknowledgements

This study was supported by the Ministry for Primary Industries, Massey University School of Veterinary Science and the Biological Heritage National Science Challenge. The authors would also like to appreciate all the support provided by the staff and students at Massey University's EpiCentre.

2.9. References

- Aandahl RZ, Reyes JF, Sisson SA, Tanaka MM. A model-based Bayesian estimation of the rate of evolution of VNTR loci in *Mycobacterium tuberculosis*. *PLoS Comput. Biol.* 8, e1002573, 2012
- Acuna-Hidalgo R, Veltman JA, Hoischen A. New insights into the generation and role of *de novo* mutations in health and disease. *Genome Biol.* 17, 241-250, 2016
- Ahlstrom C, Barkema HW, Stevenson K, Zadoks RN, Biek R, Kao R, Trewby H, Hauptstein D, Kelton DF, Fecteau G, Labrecque O, Keefe GP, McKenna SLB, De Buck J. Limitations of variable number of tandem repeat typing identified through whole genome sequencing of *Mycobacterium avium* subsp. paratuberculosis on a national and herd level. *BMC Genom.* 16, 161-169, 2015
- Aitken CK, McCaw RF, Bowden DS, Tracy SL, Kelsall JG, Higgs PG, Kerger MJ, Nguyen H, Crofts JN. Molecular epidemiology of hepatitis C virus in a social network of injection drug users. *J. Infect. Dis.* 190, 1586-95, 2004
- Ajith Kumar AK, Anoop Kumar AS. Deadly Nipah Outbreak in Kerala: Lessons Learned for the Future. *Indian J. Crit. Care Med.* 22, 475-476, 2018
- Aldous DJ. Stochastic Models and Descriptive Statistics for Phylogenetic Trees, from Yule to Today. *Statist. Sci.* 16, 23-34, 2001
- Alizon S, Hurford A, Mideo N, Baalen MV. Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *J. Evol. Biol.* 22, 245-259, 2009
- Alkhamis MA, Arruda AG, Morrison RB, Perez AM. Novel approaches for Spatial and Molecular Surveillance of Porcine Reproductive and Respiratory Syndrome Virus (PRRSv) in the United States. *Sci. Rep.* 7, 4343, 2017
- Ames GM, George DB, Hampson CP, Kanarek AR, McBee CD, Lockwood DR, Achter JD, Webb CT. Using network properties to predict disease dynamics on human contact networks. *Proc. Biol. Sci.* 278, 3544-50, 2011
- Anderson RM, May RM. Population biology of infectious diseases: Part I. *Nature* 280, 361-7, 1979
- Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB. Estimating Divergence Times from Molecular Data on Phylogenetic and Population Genetic Timescales. *Ann. Rev. Ecol. Syst.* 33, 707-740, 2002

- Ashkenazy H, Penn O, Doron-Faigenboim A, Cohen O, Cannarozzi G, Zomer O, Pupko T. FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res.* 40, 580-584, 2012
- Bansal S, Grenfell BT, Meyers LA. When individual behaviour matters: homogeneous and network models in epidemiology. *J. R. Soc. Interface.* 4, 879-91, 2007
- Bansal S, Read J, Pourbohloul B, Meyers LA. The dynamic nature of contact networks in infectious disease epidemiology. *J. Biol. Dyn.* 4, 478-89, 2010
- Barabási AL, Albert R. Emergence of scaling in random networks. *Science* 286, 509-512, 1999
- Barbosa MT, Byington MR, Struchiner CJ. Dynamic models and social networks: a review and reflections on their contribution to understanding the HIV epidemic. *Cad. Saude. Publica.* 16, 37-51, 2000
- Barrett LG, Thrall PH, Burdon JJ, Nicotra AB, Linde CC. Population structure and diversity in sexual and asexual populations of the pathogenic fungus *Melampsora lini*. *Mol. Ecol.* 17, 3401-15, 2008
- Bataille A, van der Meer F, Stegeman A, Koch G. Evolutionary Analysis of Inter-Farm Transmission Dynamics in a Highly Pathogenic Avian Influenza Epidemic. *PLoS Pathog.* 7, e1002094, 2011
- Bell DC, Montoya ID, Atkinson JS, Yang SJ. Social networks and forecasting the spread of HIV infection. *J. Acquir. Immune Defic. Syndr.* 31, 218-29, 2002
- Bergholz PW, Noar JD, Buckley DH. Environmental patterns are imposed on the population structure of *Escherichia coli* after fecal deposition. *Appl. Environ. Microbiol.* 77, 211-9, 2011
- Biek R, Pybus OG, Lloyd-Smith JO, Didelot X. Measurably evolving pathogens in the genomic era. *Trends Ecol. Evol. (Amst.)* 30, 306-313, 2015
- Bielejec F, Lemey P, Baele G, Rambaut A, Suchard MA. Inferring heterogeneous evolutionary processes through time: from sequence substitution to phylogeography. *Syst. Biol.* 63, 493-504, 2014
- Bos DH, Posada D. Using models of nucleotide evolution to build phylogenetic trees. *Developm. Comparat. Immunol.* 29, 211-227, 2005

- Booth RE, Thomas CJ, El-Attar LM, Gunn G, Brownlie J.** A phylogenetic analysis of Bovine Viral Diarrhoea Virus (BVDV) isolates from six different regions of the UK and links to animal movement data. *Vet. Res.* 44, 43, 2013
- Boots M, Hudson PJ, Sasaki A.** Large shifts in pathogen virulence relate to host population structure. *Science* 303, 842-4, 2004
- Brauer F.** Mathematical epidemiology: Past, present, and future. *Infect. Dis. Model.* 2, 113–127, 2017
- Brewer DD, Hagan H, Sullivan DG, Muth SQ, Hough ES, Feuerborn NA, Gretch DR.** Social structural and behavioral underpinnings of hyperendemic hepatitis C virus transmission in drug injectors. *J. Infect. Dis.* 194, 764-72, 2006
- Briand F-X, Schmitz A, Ogor K, Le Prioux A, Guillou-Cloarec C, Guillemoto C, Allée C, Le Bras M-O, Hirchaud E, Quenault H, Touzain F, Cherbonnel-Pansart M, Lemaitre E, Courtillon C, Gares H, Daniel P, Fediaevsky A, Massin P, Blanchard Y, Eterradosi N, van der Werf S, Jestin V, Niqueux E.** Emerging highly pathogenic H5 avian influenza viruses in France during winter 2015/16: phylogenetic analyses and markers for zoonotic potential. *Euro Surveill.* 22, 30473, 2017
- Brown E, Dessai U, McGarry S, Gerner-Smidt P.** Use of Whole-Genome Sequencing for Food Safety and Public Health in the United States. *Foodborne Pathog. Dis.* 16, 441-50, 2019
- Buckee CO, Koelle K, Mustard MJ, Gupta S.** The effects of host contact network structure on pathogen diversity and strain structure. *Proc. Natl. Acad. Sci. USA* 101, 10839-44, 2004
- Campbell F, Strang C, Ferguson N, Cori A, Jombart T.** When are pathogen genome sequences informative of transmission events? *PLoS Pathog.* 14, e1006885, 2018
- Carrel MA, Emch M, Nguyen T, Todd Jobe R, Wan XF.** Population-environment drivers of H5N1 avian influenza molecular change in Vietnam. *Health Place* 18, 1122-31, 2012
- Carrel M, Patel J, Taylor SM, Janko M, Mwandagalirwa MK, Tshetu AK, Escalante AA, McCollum A, Alam MT, Udhayakumar V, Meshnick S, Emch M.** The geography of malaria genetics in the Democratic Republic of Congo: A complex and fragmented landscape. *Soc. Sci. Med.* 133, 233-41, 2015

- Carrique-Mas JJ, Medley GF, Green LE.** Risks for bovine tuberculosis in British cattle farms restocked after the foot and mouth disease epidemic of 2001. *Prev. Vet. Med.* 84, 85-93, 2008
- Choi KH, Ning Z, Gregorich SE, Pan QC.** The influence of social and sexual networks in the spread of HIV and syphilis among men who have sex with men in Shanghai, China. *J. Acquir. Immune Defic. Syndr.* 45, 77-84, 2007
- Christley RM, French NP.** Small-world topology of UK racing: the potential for rapid spread of infectious agents. *Equine Vet. J.* 35, 586-9, 2003
- Christley RM, Pinchbeck GL, Bowers RG, Clancy D, French NP, Bennett R, Turner J.** Infection in social networks: using network analysis to identify high-risk individuals. *Am. J. Epidemiol.* 162, 1024-31, 2005
- Christley RM, Robinson SE, Moore B, Setzkorn C, Donald I.** Responses of farmers to introduction in England and Wales of pre-movement testing for bovine tuberculosis. *Prev. Vet. Med.* 100, 126-33, 2011
- Colijn C, Gardy J.** Phylogenetic tree shapes resolve disease transmission patterns. *Evol. Med. Public Health* 2014, 96-108, 2014
- Cook VJ, Sun SJ, Tapia J, Muth SQ, Arguello DF, Lewis BL, Rothenberg RB, McElroy PD.** Transmission network analysis in tuberculosis contact investigations. *J. Infect. Dis.* 196, 1517-27, 2007
- Corner LA, Pfeiffer DU, Morris RS.** Social-network analysis of *Mycobacterium bovis* transmission among captive brushtail possums (*Trichosurus vulpecula*). *Prev. Vet. Med.* 59, 147-67, 2003
- Cottam EM, Thebaud G, Wadsworth J, Gloster J, Mansley L, Paton DJ, King DP, Haydon DT.** Integrating genetic and epidemiological data to determine transmission pathways of foot-and-mouth disease virus. *Proc. Biol. Sci.* 275, 887-95, 2008
- Craft ME.** Infectious disease transmission and contact networks in wildlife and livestock. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 2015
- Cross PC, Creech TG, Ebinger MR, Heisey DM, Irvine KM, Creel S.** Wildlife contact analysis: emerging methods, questions, and challenges. *Behav. Ecol. Sociobiol.* 66, 1437-1447, 2012
- Croucher NJ, Didelot X.** The application of genomics to tracing bacterial pathogen transmission. *Curr. Opin. Microbiol.* 23, 62-7, 2015

- Curtis R, Friedman SR, Neaigus A, Jose B, Goldstein M, Ildefonso G. Street-level drug markets: network structure and HIV risk. *Social Networks* 17, 229-49, 1995
- D'Angelo-Scott H, Cutler J, Friedman D, Hendriks A, Jolly AM. Social network investigation of a syphilis outbreak in Ottawa, Ontario. *Can. J. Infect. Dis. Med. Microbiol.* 26, 268-72, 2015
- Danon L, Ford AP, House T, Jewell CP, Keeling MJ, Roberts GO, Ross JV, Vernon MC. Networks and the epidemiology of infectious disease. *Interdiscip. Perspect. Infect. Dis.* 2011, 284909, 2011
- De P, Singh AE, Wong T, Yacoub W, Jolly AM. Sexual network analysis of a gonorrhoea outbreak. *Sex. Transm. Infect.* 80, 280-5, 2004
- de Angelis D, Presanis AM, Birrell PJ, Tomba GS, House T. Four key challenges in infectious disease modelling using data from multiple sources. *Epidemics* 10, 83-7, 2015
- de Castro F, Bolker B. Mechanisms of disease-induced extinction. *Ecology Letters* 8, 117-26, 2005
- Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the Basic Reproduction Number (R_0). *Emerg. Infect. Dis.* 25, 1-4, 2019
- de Maio ND, Wu C-H, Wilson DJ. SCOTTI: Efficient Reconstruction of Transmission within Outbreaks with the Structured Coalescent. *PLoS Comput. Biol.* 12, e1005130, 2016
- de Smet R, Marchal K. Advantages and limitations of current network inference methods. *Nature Rev. Microbiol.* 8, 717, 2010
- Deng X, Bakker HCd, Hendriksen RS. Genomic Epidemiology: Whole-Genome-Sequencing-Powered Surveillance and Outbreak Investigation of Foodborne Bacterial Pathogens. *Ann. Rev. Food Sci. Tech.* 7, 353-74, 2016
- Dent JE, Kao RR, Kiss IZ, Hyder K, Arnold M. Contact structures in the poultry industry in Great Britain: exploring transmission routes for a potential avian influenza virus epidemic. *BMC Vet. Res.* 4, 27, 2008

- Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW.** Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genet.* 13, 601–612, 2012
- Didelot X, Walker S, Peto TE, Crook DW, Wilson DJ.** Within-host evolution of bacterial pathogens. *Nat. Rev. Microbiol.* 14, 150–62, 2016
- Diniz-Filho JA, Diniz JV, Rangel TF, Soares TN, Telles MP, Collevatti RG, Bini LM.** A new eigenfunction spatial analysis describing population genetic structure. *Genetica* 141, 479–89, 2013
- Douady CJ, Delsuc F, Boucher Y, Doolittle WF, Douzery EJP.** Comparison of Bayesian and Maximum Likelihood Bootstrap Measures of Phylogenetic Reliability. *Mol. Biol. Evol.* 20, 248–254, 2003
- Drummond AJ, Rambaut A.** BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214, 2007
- Dube C, Ribble C, Kelton D, McNab B.** A review of network analysis terminology and its application to foot-and-mouth disease modelling and policy development. *Transbound. Emerg. Dis.* 56, 73–85, 2009
- Dunn JR.** Whole-Genome Sequencing: Opportunities and Challenges for Public Health, Food-borne Outbreak Investigations, and the Global Food Supply. *J. Infect. Dis.* 213, 499–501, 2016
- Eames K, Bansal S, Frost S, Riley S.** Six challenges in measuring contact networks for use in modelling. *Epidemics* 10, 72–77, 2015
- Eames KTD, Keeling MJ.** Contact tracing and disease control. *Proc. R. Soc. Lond. B Biol. Sci.* 270, 2565–2571, 2003
- Ebert D, Bull JJ.** Challenging the trade-off model for the evolution of virulence: is virulence management feasible? *Trends Microbiol.* 11, 15–20, 2003
- Edwards SV.** Natural selection and phylogenetic analysis. *Proc. Natl. Acad. Sci. USA* 106, 8799–800, 2009
- El-Sayed AM, Scarborough P, Seemann L, Galea S.** Social network analysis and agent-based modeling in social epidemiology. *Epidemiol. Perspect. Innov.* 9, 1–9, 2012
- Enright J, Kao RR.** Epidemics on dynamic networks. *Epidemics* 24, 88–97, 2018

Ewald PW. The evolution of virulence. *Sci. Am.* 268, 86-93, 1993

Eydboosh S, Haghdooost AA, Mostafavi E, Bahrapour A, Azadmanesh K, Zolala F. Molecular epidemiology of infectious diseases. *Electron Physicians* 9, 5149-58, 2017

Famulare M, Hu H. Extracting transmission networks from phylogeographic data for epidemic and endemic diseases: Ebola virus in Sierra Leone, 2009 H1N1 pandemic influenza and polio in Nigeria. *Int. Health* 7, 130–138, 2015

Farine DR, Whitehead H. Constructing, conducting and interpreting animal social network analysis. *J. Ani. Ecol.* 84, 1144–1163, 2015

Fefferman NH, Ng KL. How disease models in static networks can fail to approximate disease in dynamic networks. *Phys. Rev.* 76, e031919, 2007

Felsenstein J. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* 17, 368–376, 1981

Field N, Cohen T, Struelens MJ, Palm D, Cookson B, Glynn JR, Gallo V, Ramsay M, Sonnenberg P, Maccannell D, Charlett A, Egger M, Green J, Vineis P, Abubakar I. Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID): an extension of the STROBE statement. *Lancet Infect. Dis.* 14, 341–352, 2014

Firestone SM, Hayama Y, Bradhurst R, Yamamoto T, Tsutsui T, Stevenson MA. Reconstructing foot-and-mouth disease outbreaks: a methods comparison of transmission network models. *Sci. Rep.* 9, 1–12, 2019

Ford CB, Lin PL, Chase MR, Shah RR, Iartchouk O, Galagan J, Mohaideen N, Ioerger TR, Sacchettini JC, Lipsitch M, Flynn JL, Fortune SM. Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nat. Genet.* 43, 482–486, 2011

Fourment M, Holmes EC. Novel non-parametric models to estimate evolutionary rates and divergence times from heterochronous sequence data. *BMC Evol. Biol.* 14, 163, 2014

Fournie G, Guitian J, Desvaux S, Cuong VC, Dung do H, Pfeiffer DU, Mangtani P, Ghani AC. Interventions for avian influenza A (H5N1) risk management in live bird market networks. *Proc. Natl. Acad. Sci. USA* 110, 9177-82, 2013

- Freeman LC.** Centrality in social networks conceptual clarification. *Social Networks* 1, 215-39, 1978
- French N, Barrigas M, Brown P, Ribiero P, Williams N, Leatherbarrow H, Birtles R, Bolton E, Fearnhead P, Fox A.** Spatial epidemiology and natural population structure of *Campylobacter jejuni* colonizing a farmland ecosystem. *Environ. Microbiol.* 7, 1116-26, 2005
- Friedman SR, Aral S.** Social Networks, Risk-Potential Networks, Health, and Disease. *J. Urban Health* 78, 411–418, 2001
- Frost SD, Pybus OG, Gog JR, Viboud C, Bonhoeffer S, Bedford T.** Eight challenges in phylodynamic inference. *Epidemics* 10, 88-92, 2015
- Funk S, Salathe M, Jansen VA.** Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J. R. Soc. Interface* 7, 1247-56, 2010
- Garbelotto M, Guglielmo F, Mascheretti S, Croucher PJ, Gonthier P.** Population genetic analyses provide insights on the introduction pathway and spread patterns of the North American forest pathogen *Heterobasidion irregulare* in Italy. *Mol. Ecol.* 22, 4855-69, 2013
- Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodtkin E, Rempel S, Moore R, Zhao Y, Holt R, Varhol R, Birol I, Lem M, Sharma MK, Elwood K, Jones SJ, Brinkman FS, Brunham RC, Tang P.** Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N. Engl. J. Med.* 364, 730-9, 2011
- Gates MC, Volkova VV, Woolhouse ME.** Risk factors for bovine tuberculosis in low incidence regions related to the movements of cattle. *BMC Vet. Res.* 9, 225-231, 2013
- Gates MC, Woolhouse ME.** Controlling infectious disease through the targeted manipulation of contact network structure. *Epidemics* 12, 11-9, 2015
- Gavryushkina A, Welch D, Stadler T, Drummond AJ.** Bayesian Inference of Sampled Ancestor Trees for Epidemiology and Fossil Calibration. *PLoS Comput. Biol.* 10, e1003919, 2014
- Gilbert M, Mitchell A, Bourn D, Mawdsley J, Clifton-Hadley R, Wint W.** Cattle movements and bovine tuberculosis in Great Britain. *Nature* 435, 491-6, 2005

- Gogarten JP, Townsend JP.** Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* 3, 679-87, 2005
- Gonzalez-Candelas F, Bracho MA, Moya A.** Molecular epidemiology and forensic genetics: application to a hepatitis C virus transmission event at a hemodialysis unit. *J. Infect. Dis.* 187, 352-8, 2003
- Gottdenker NL, Streicker DG, Faust CL, Carroll CR.** Anthropogenic Land Use Change and Infectious Diseases: A Review of the Evidence. *EcoHealth* 11, 619–632, 2014
- Grange ZL, Gartrell BD, Biggs PJ, Nelson NJ, Anderson M, French NP.** Microbial Genomics of a Host-Associated Commensal Bacterium in Fragmented Populations of Endangered Takahe. *Microb. Ecol.* 71, 1020-9, 2016
- Grant H, Foss AM, Watts C, Medley GF, Mukandavire Z.** Is modelling complexity always needed? Insights from modelling PrEP introduction in South Africa. *J. Public Health (Oxf.)* fdz178, 2020
- Grear DA, Kaneene JB, Averill JJ, Webb CT.** Local cattle movements in response to ongoing bovine tuberculosis zonation and regulations in Michigan, USA. *Prev. Vet. Med.* 114, 201-12, 2014
- Green DM, Kiss IZ, Kao RR.** Modelling the initial spread of foot-and-mouth disease through animal movements. *Proc. Biol. Sci.* 273, 2729-35, 2006
- Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, Holmes EC.** Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303, 327-32, 2004
- Handcock RN, Swain DL, Bishop-Hurley GJ, Patison KP, Wark T, Valencia P, Corke P, O'Neill CJ.** Monitoring Animal Behaviour and Environmental Interactions Using Wireless Sensor Networks, GPS Collars and Satellite Remote Sensing. *Sensors* 9, 3586–3603, 2009
- Harrison CJ, Langdale JA.** A step by step guide to phylogeny reconstruction. *The Plant J.* 45, 561–572, 2006
- Hasegawa M, Kishino H, Yano T.** Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22, 160–174, 1985

- He M, Miyajima F, Roberts P, Ellison L, Pickard DJ, Martin MJ, Connor TR, Harris SR, Fairley D, Bamford KB, D'Arc S, Brazier J, Brown D, Coia JE, Douce G, Gerding D, Kim HJ, Koh TH, Kato H, Senoh M, Louie T, Michell S, Butt E, Peacock SJ, Brown NM, Riley T, Songer G, Wilcox M, Pirmohamed M, Kuijper E, Hawkey P, Wren BW, Dougan G, Parkhill J, Lawley TD. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nat. Genet.* 45, 109–113, 2013
- Hendry AP, Gotanda KM, Svensson EI. Human influences on evolution, and the ecological and societal consequences. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372, e20160028, 2017
- Hidano A, Gates MC. Assessing biases in phylodynamic inferences in the presence of super-spreaders. *Vet. Res.* 50, 74–82, 2019
- Ho SYW, Larson G. Molecular clocks: when times are a-changin'. *Trends Genet.* 22, 79–83, 2006
- Holder M, Lewis PO. Phylogeny estimation: traditional and Bayesian approaches. *Nat. Rev. Genet.* 4, 275–284, 2003
- Holmes EC, Grenfell BT. Discovering the Phylodynamics of RNA Viruses. *PLoS Comput. Biol.* 5, e1000505, 2009
- Holme P, Kim BJ. Growing scale-free networks with tunable clustering. *Phys. Rev.* 65–71, 2002
- Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755, 2001
- Inns T, Ashton PM, Herrera-Leon S, Lighthill J, Foulkes S, Jombart T, Rehman Y, Fox A, Dallman T, EDEP, Browning L, Coia JE, Edeghere O, Vivancos R. Prospective use of whole genome sequencing (WGS) detected a multi-country outbreak of *Salmonella* Enteritidis. *Epidemiol. Infect.* 145, 289–98, 2017
- Jackson RW, Johnson LJ, Clarke SR, Arnold DL. Bacterial pathogen evolution: breaking news. *Trends Genet.* 21, 32–40, 2011
- Jaros P, Cookson AL, Campbell DM, Besser TE, Shringi S, Mackereth GF, Lim E, Lopez L, Dufour M, Marshall JC, Baker MG, Hathaway S, Prattley DJ, French NP. A prospective case-control and molecular epidemiological study of human cases of Shiga toxin-producing *Escherichia coli* in New Zealand. *BMC Infect. Dis.* 13, 450, 2013

- Jolad S, Liu W, Schmittmann B, Zia RK. Epidemic spreading on preferred degree adaptive networks. *PLoS One* 7, e48686, 2012
- Joly S, Bryant D, Lockhart PJ. Flexible methods for estimating genetic distances from single nucleotide polymorphisms. *Meth. Ecol. Evol.* 6, 938–948, 2015
- Jombart T, Aanensen DM, Baguelin M, Birrell P, Cauchemez S, Camacho A, Colijn C, Collins C, Cori A, Didelot X, Fraser C, Frost S, Hens N, Hugues J, Höhle M, Opatowski L, Rambaut A, Ratmann O, Soubeyrand S, Suchard MA, Wallinga J, Ypma R, Ferguson N. OutbreakTools: a new platform for disease outbreak analysis using the R software. *Epidemics* 7, 28–34, 2014a
- Jombart T, Eggo RM, Dodd PJ, Balloux F. Reconstructing disease outbreaks from genetic data: a graph approach. *Heredity (Edinb.)* 106, 383–90, 2011
- Jombart T, Cori A, Didelot X, Cauchemez S, Fraser C, Ferguson N. Bayesian reconstruction of disease outbreaks by combining epidemiologic and genomic data. *PLoS Comput. Biol.* 10, e1003457, 2014b
- Jukes TH, Cantor CR. *Evolution of Protein Molecules*. In: Mammalian Protein Metabolism. Academic Press, New York, 1969
- Kamath PL, Foster JT, Drees KP, Luikart G, Quance C, Anderson NJ, Clarke PR, Cole EK, Drew ML, Edwards WH, Rhyan JC, Treanor JJ, Wallen RL, White PJ, Robbe-Austerman S, Cross PC. Genomics reveals historic and contemporary transmission dynamics of a bacterial disease among wildlife and livestock. *Nat. Commun.* 7, e11448, 2016
- Kao RR. Evolution of pathogens towards low R_0 in heterogeneous populations. *J. Theor. Biol.* 242, 634–42, 2006
- Kao RR, Green DM, Johnson J, Kiss IZ. Disease dynamics over very different timescales: foot-and-mouth disease and scrapie on the network of livestock movements in the UK. *J. R. Soc. Interface* 4, 907–16, 2007
- Keeling M. Evolutionary trade-offs at two timescales: competition versus persistence. *Proc. Biol. Sci.* 267, 385–91, 2000
- Keeling M. The implications of network structure for epidemic dynamics. *Theor. Popul. Biol.* 67, 1–8, 2005

- Keeling MJ, Eames KT. Networks and epidemic models. *J. R. Soc. Interface* 2, 295-307, 2005
- Keeling M, Rohani P. *Modeling infectious diseases in humans and animals*. Princeton University Press, Princeton, New Jersey, 2008
- Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111-120, 1980
- Kingman JF. Origins of the coalescent. 1974-1982. *Genetics* 156, 1461-3, 2000
- Kiss IZ, Green DM, Kao RR. Infectious disease control using contact tracing in random and scale-free networks. *J. R. Soc. Interface* 3, 55-62, 2006a
- Kiss IZ, Green DM, Kao RR. The effect of contact heterogeneity and multiple routes of transmission on final epidemic size. *Math. Biosci.* 203, 124-36, 2006b
- Kiss IZ, Green DM, Kao RR. The effect of network mixing patterns on epidemic dynamics and the efficacy of disease contact tracing. *J. R. Soc. Interface* 5, 791-9, 2008
- Klovdahl AS, Graviss EA, Yaganehdoost A, Ross MW, Wanger A, Adams GJ, Musser JM. Networks and tuberculosis: an undetected community outbreak involving public places. *Soc. Sci. Med.* 52, 681-94, 2001
- Köser CU, Holden MTG, Ellington MJ, Cartwright EJP, Brown NM, Ogilvy-Stuart AL, Hsu LY, Chewapreecha C, Croucher NJ, Harris SR, Sanders M, Enright MC, Dougan G, Bentley SD, Parkhill J, Fraser LJ, Betley JR, Schulz-Trieglaff OB, Smith GP, Peacock SJ. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N. Engl. J. Med.* 366, 2267-2275, 2012
- Kossinets G. Effects of missing data in social networks. *Social Networks* 28, 247-68, 2006
- Kouyos RD, von Wyl V, Yerly S, Böni J, Taffé P, Shah C, Bürgisser P, Klimkait T, Weber R, Hirschel B, Cavassini M, Furrer H, Battegay M, Vernazza PL, Bernasconi E, Rickenbach M, Ledergerber B, Bonhoeffer S, Günthard HF. Molecular Epidemiology Reveals Long-Term Changes in HIV Type 1 Subtype B Transmission in Switzerland. *J. Infect. Dis.* 201, 1488-1497, 2010
- Kretzschmar M, Gomes MGM, Coutinho RA, Koopman JS. Unlocking pathogen genotyping information for public health by mathematical modeling. *Trends Microbiol.* 18, 406-412, 2010

- Kühnert D, Wu C-H, Drummond AJ.** Phylogenetic and epidemic modelling of rapidly evolving infectious diseases. *Infect. Genet. Evol.* 11, 1825–1841, 2011
- Kumar S.** Molecular clocks: four decades of evolution. *Nat. Rev. Genet.* 6, 654–662, 2005
- Lau MSY, Marion G, Streftaris G, Gibson G.** A Systematic Bayesian Integration of Epidemiological and Genetic Data. *PLoS Comput. Biol.* 11, e1004633, 2015
- Lebarbenchon C, Brown SP, Poulin R, Gauthier-Clerc M, Thomas F.** Evolution of pathogens in a man-made world. *Mol. Ecol.* 17, 475–84, 2008
- Lee K, Brumme ZL.** Operationalizing the One Health approach: the global governance challenges. *Health Policy Plan* 28, 778–785, 2013
- Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, Russell CA, Smith DJ, Pybus OG, Brockmann D, Suchard MA.** Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS Pathog.* 10, e1003932, 2014
- Lemey P, Rambaut A, Drummond AJ, Suchard MA.** Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* 5, e1000520, 2009
- Lessler J, Cummings DAT.** Mechanistic Models of Infectious Disease and Their Impact on Public Health. *Am. J. Epidemiol.* 183, 415–422, 2016
- Leventhal GE, Hill AL, Nowak MA, Bonhoeffer S.** Evolution and emergence of infectious diseases in theoretical and real-world networks. *Nat. Comm.* 6, e6101, 2015
- Levin BR, Bull JJ.** Short-sighted evolution and the virulence of pathogenic microorganisms. *Trends Microbiol.* 2, 76–81, 1994
- Lewis NS, Verhagen JH, Javakhishvili Z, Russell CA, Lexmond P, Westgeest KB, Bestebroer TM, Halpin RA, Lin X, Ransier A, Fedorova NB, Stockwell TB, Latorre-Margalef N, Olsen B, Smith G, Bahl J, Wentworth DE, Waldenström J, Fouchier RAM, de Graaf M.** Influenza A virus evolution and spatio-temporal dynamics in Eurasian wild birds: a phylogenetic and phylogeographical study of whole-genome sequence data. *J. Gen. Virol.* 96, 2050–60, 2015
- Liang L, Xu B, Chen Y, Liu Y, Cao W, Fang L, Feng L, Goodchild MF, Gong P.** Combining Spatial-Temporal and Phylogenetic Analysis Approaches for Improved Understanding on Global H5N1 Transmission. *PLoS One* 5, e13575, 2010

- Lin CJ, Deger KA, Tien JH.** Modeling the trade-off between transmissibility and contact in infectious disease dynamics. *Math. Biosci.* 277, 15-24, 2016
- Liu Q-H, Ajelli M, Aleta A, Merler S, Moreno Y, Vespignani A.** Measurability of the epidemic reproduction number in data-driven contact networks. *PNAS* 115, 12680–12685, 2018
- Lively CM.** Migration, Virulence, and the Geographic Mosaic of Adaptation by Parasites. *The Am. Nat.* 153, 34-47, 1999
- Luke DA, Harris JK.** Network Analysis in Public Health: History, Methods, and Applications. *Ann. Rev. Public Health* 28, 69–93, 2007
- Maharjan R, Ferenci T.** Mutational signatures indicative of environmental stress in bacteria. *Mol. Biol. Evol.* 32, 380–391, 2015
- Malave A, Elamin EM.** Severe Acute Respiratory Syndrome (SARS): Lessons for Future Pandemics. *AMA J. Ethics* 12, 719–725, 2010
- Marceau V, Noel PA, Hebert-Dufresne L, Allard A, Dube LJ.** Adaptive networks: Coevolution of disease and topology. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 82, e036116, 2010
- Marquetoux N, Heuer C, Wilson P, Ridler A, Stevenson M.** Merging DNA typing and network analysis to assess the transmission of paratuberculosis between farms. *Prev. Vet. Med.* 134, 113-21, 2016
- Martin V, Zhou X, Marshall E, Jia B, Fusheng G, FrancoDixon MA, DeHaan N, Pfeiffer DU, Soares Magalhaes RJ, Gilbert M.** Risk-based surveillance for avian influenza control along poultry market chains in South China: The value of social network analysis. *Prev. Vet. Med.* 102, 196-205, 2011
- Martinez-Lopez B, Perez AM, Sanchez-Vizcaino JM.** Social network analysis. Review of general concepts and use in preventive veterinary medicine. *Transbound. Emerg. Dis.* 56, 109-20, 2009
- Masuda N, Holme P.** Predicting and controlling infectious disease epidemics using temporal networks. *Prime Rep.* 5, 6-14, 2013
- McDonald BA, Linde C.** Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* 40, 349-79, 2002

- McElroy PD, Rothenberg RB, Varghese R, Woodruff R, Minns GO, Muth SQ, Lambert LA, Ridzon R.** A network-informed approach to investigating a tuberculosis outbreak: implications for enhancing contact investigations. *Int. J. Tuberc. Lung Dis.* 7, 486-93, 2003
- Messinger SM, Ostling A.** The consequences of spatial structure for the evolution of pathogen transmission rate and virulence. *Am. Nat.* 174, 441–454, 2009
- Metcalfe CJ, Birger RB, Funk S, Kouyos RD, Lloyd-Smith JO, Jansen VA.** Five challenges in evolution and infectious diseases. *Epidemics* 10, 40-4, 2015
- Meyers LA, Pourbohloul B, Newman ME, Skowronski DM, Brunham RC.** Network theory and SARS: predicting outbreak diversity. *J. Theor. Biol.* 232, 71-81, 2005
- Mikolajczyk RT, Kretzschmar M.** Collecting social contact data in the context of disease transmission: Prospective and retrospective study designs. *Social Networks* 30, 127–135, 2008
- Morris M, Zavisca J, Dean L.** Social and sexual networks: their role in the spread of HIV/AIDS among young gay men. *AIDS Educ. Prev.* 7, 24-35, 1995
- Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J, Sadkowska-Todys M, Rosinska M, Edmunds WJ.** Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* 5, 74-85, 2008
- Mullner P, Collins-Emerson JM, Midwinter AC, Carter P, Spencer SE, van der Logt P, Hathaway S, French NP.** Molecular epidemiology of *Campylobacter jejuni* in a geographically isolated country with a uniquely structured poultry industry. *Appl. Environ. Microbiol.* 76, 2145-54, 2010
- Munster VJ, Bausch DG, de Wit E, Fischer R, Kobinger G, Muñoz-Fontela C, Olson SH, Seifert SN, Sprecher A, Ntoumi F, Massaquoi M, Mombouli J-V.** Outbreaks in a Rapidly Changing Central Africa — Lessons from Ebola. *N. Eng. J. M.* 379, 1198-201, 2018
- Nadon C, van Walle I, Gerner-Smidt P, Campos J, Chinen I, Concepcion-Acevedo J, Gilpin B, Smith AM, Man Kam K, Perez E, Trees E, Kubota K, Takkinen J, Nielsen EM, Carleton H, Panel F-NE.** PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. *Euro Surveill.* 22, e30544, 2017

- Newman ME.** Assortative mixing in networks. *Phys. Rev. Lett.* 89, e208701, 2002
- Newman ME.** Properties of highly clustered networks. *Phys. Rev.* e68, 2003
- Ochman H, Lawrence JG, Groisman EA.** Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299-304, 2000
- O'Malley AJ, Marsden PV.** The Analysis of Social Networks. *Health Serv. Outcomes Res. Methodol.* 8, 222–269, 2008
- Omedo I, Mogeni P, Rockett K, Kamau A, Hubbart C, Jeffreys A, Ochola-Oyier LI, de Villiers EP, Gitonga CW, Noor AM, Snow RW, Kwiatkowski D, Bejon P.** Geographic-genetic analysis of *Plasmodium falciparum* parasite populations from surveys of primary school children in Western Kenya. *Wellcome Open Res.* 2, 29, 2017
- Ortiz-Pelaez A, Pfeiffer DU, Soares-Magalhaes RJ, Guitian FJ.** Use of social network analysis to characterize the pattern of animal movements in the initial phases of the 2001 foot and mouth disease (FMD) epidemic in the UK. *Prev. Vet. Med.* 76, 40-55, 2006
- Osmani MG, Ward MP, Giasuddin M, Islam MR, Kalam A.** The spread of highly pathogenic avian influenza (subtype H5N1) clades in Bangladesh, 2010 and 2011. *Prev. Vet. Med.* 114, 21–27, 2014
- Otte E, Rousseau R.** Social network analysis: a powerful strategy, also for the information sciences. *J. Inform. Sci.* 28, 441–453, 2002
- Pagel M.** Inferring the historical patterns of biological evolution. *Nature* 401, 877–884, 1999
- Parham PE, Ferguson NM.** Space and contact networks: capturing the locality of disease transmission. *J. R. Soc. Interface* 3, 483-93, 2006
- Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A.** Epidemic processes in complex networks. *Rev. Mod. Phys.* 87, 925–979, 2015
- Pfeiffer CN, Firestone SM, Campbell AJD, Larsen JWA, Stevenson MA.** Sample size considerations for livestock movement network data. *Prev. Vet. Med.* 122, 399–405, 2015

- Pightling AW, Pettengill JB, Luo Y, Baugher JD, Rand H, Strain E.** Interpreting Whole-Genome Sequence Analyses of Foodborne Bacteria for Regulatory Applications and Outbreak Investigations. *Front. Microbiol.* 9, e1482, 2018
- Pilon R, Leonard L, Kim J, Vallee D, De Rubeis E, Jolly AM, Wylie J, Pelude L, Sandstrom P.** Transmission patterns of HIV and hepatitis C virus among networks of people who inject drugs. *PLoS One* 6, e22245, 2011
- Poolkhet C, Chairatanayuth P, Thongratsakul S, Kasemsuwan S, Rukkwamsuk T.** Social network analysis used to assess the relationship between the spread of avian influenza and movement patterns of backyard chickens in Ratchaburi, Thailand. *Res. Vet. Sci.* 95, 82-6, 2013
- Prado F, Sheih A, West JD, Kerr B.** Coevolutionary cycling of host sociality and pathogen virulence in contact networks. *J. Theor. Biol.* 261, 561-9, 2009
- Price JR, Golubchik T, Cole K, Wilson DJ, Crook DW, Thwaites GE, Bowden R, Walker AS, Peto TEA, Paul J, Llewelyn MJ.** Whole-genome sequencing shows that patient-to-patient transmission rarely accounts for acquisition of *Staphylococcus aureus* in an intensive care unit. *Clin. Infect. Dis.* 58, 609–618, 2014
- Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH.** The epidemic behaviour of the hepatitis C virus. *Science* 292, 2323-5, 2001
- Ranjbar R, Karami A, Farshad S, Giammanco GM, Mammina C.** Typing methods used in the molecular epidemiology of microbial pathogens: a how-to guide. *New Microbiol.* 37, 1–15, 2014
- Rannala B, Yang Z.** Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. *J. Mol. Evol.* 43, 304–311, 1996
- Read JM, Keeling MJ.** Disease evolution on networks: the role of contact structure. *Proc. Biol. Sci.* 270, 699-708, 2003
- Read JM, Eames KT, Edmunds WJ.** Dynamic social networks and the implications for the spread of infectious disease. *J. R. Soc. Interface* 5, 1001-7, 2008
- Restif O.** Evolutionary epidemiology 20 years on: Challenges and prospects. *Infect., Genet. Evol.* 9, 108–123, 2009
- Ridenhour B, Kowalik JM, Shay DK.** Unravelling R0: Considerations for Public Health Applications. *Am. J. Public Health* 104, 32–41, 2014

- Riley LW. *Molecular epidemiology of infectious diseases: principles and practices*. ASM Press, Washington, D.C. 2004
- Riley S, Fraser C, Donnelly CA, Ghani AC, Abu-Raddad LJ, Hedley AJ, Leung GM, Ho L-M, Lam T-H, Thach TQ, Chau P, Chan K-P, Lo S-V, Leung P-Y, Tsang T, Ho W, Lee K-H, Lau EMC, Ferguson NM, Anderson RM. Transmission Dynamics of the Etiological Agent of SARS in Hong Kong: Impact of Public Health Interventions. *Science* 300, 1961–1966, 2003
- Rizzo J, Rouchka EC. *Review of Phylogenetic Tree Construction*. University of Louisville Bioinformatics Laboratory Technical Report Series. University of Louisville, United States, 2007
- Robinson SE, Christley RM. Exploring the role of auction markets in cattle movements within Great Britain. *Prev. Vet. Med.* 81, 21-37, 2007
- Robinson K, Fyson N, Cohen T, Fraser C, Colijn C. How the dynamics and structure of sexual contact networks shape pathogen phylogenies. *PLoS Comput. Biol.* 9, e1003105, 2013
- Rodríguez F, Oliver JL, Marín A, Medina JR. The general stochastic model of nucleotide substitution. *J. Theor. Biol.* 142, 485–501, 1990
- Romano CM, de Carvalho-Mello IMVG, Jamal LF, de Melo FL, Iamarino A, Motoki M, Pinho JRR, Holmes EC, Zanutto PM de A. Social Networks Shape the Transmission Dynamics of Hepatitis C Virus. *PLoS One* 5, 2010
- Rosendal T, Dewey C, Friendship R, Wootton S, Young B, Poljak Z. Spatial and temporal patterns of porcine reproductive and respiratory syndrome virus (PRRSV) genotypes in Ontario, Canada, 2004-2007. *BMC Vet. Res.* 10, 83, 2014
- Ross R. An application of the theory of probabilities to the study of a priori pathometry. *R. Soc. Lond. Ser. A* 92, 204–230, 1916
- Rothenberg R, Kimbrough L, Lewis-Hardy R, Heath B, Williams OC, Tambe P, Johnson D, Schrader M. Social network methods for endemic foci of syphilis: a pilot project. *Sex. Transm. Dis.* 27, 12-8, 2000
- Savini L, Candeloro L, Conte A, De Massis F, Giovannini A. Development of a forecasting model for brucellosis spreading in the Italian cattle trade network aimed to prioritise the field interventions. *PLoS One* 12, e0177313, 2017

- Scarpino SV, Iamarino A, Wells C, Yamin D, Ndeffo-Mbah M, Wenzel NS, Fox SJ, Nyenswah T, Altice FL, Galvani AP, Meyers LA, Townsend JP.** Epidemiological and Viral Genomic Sequence Analysis of the 2014 Ebola Outbreak Reveals Clustered Transmission. *Clin. Infect. Dis.* 60, 1079–1082, 2015
- Service S, Blower SM.** HIV transmission in sexual networks: an empirical analysis. *Proc. Biol. Sci.* 260, 237–44, 1995
- Shaw LB, Schwartz IB.** Fluctuating epidemics on adaptive networks. *Physic. Rev. E* 77, e066101, 2008
- Shirley MD, Rushton SP.** Where diseases and networks collide: lessons to be learnt from a study of the 2001 foot-and-mouth disease epidemic. *Epidemiol. Infect.* 133, 1023–32, 2005a
- Shirley MDF, Rushton SP.** The impacts of network topology on disease spread. *Ecol. Complex.* 2, 287–99, 2005b
- Silk MJ, Hodgson DJ, Rozins C, Croft DP, Delahay RJ, Boots M, McDonald RA.** Integrating social behaviour, demography and disease dynamics in network models: applications to disease management in declining wildlife populations. *Philos. Trans. R Soc. B: Biol. Sci.* 374, e20180211, 2019
- Simmons MP.** A confounding effect of missing data on character conflict in maximum likelihood and Bayesian MCMC phylogenetic analyses. *Mol. Phylog. Evol.* 80, 267–280, 2014
- Smieszek T.** A mechanistic model of infection: why duration and intensity of contacts should be included in models of disease spread. *Theor. Biol. Med. Model.* 6, 25, 2009
- Smith KF, Goldberg M, Rosenthal S, Carlson L, Chen J, Chen C, Ramachandran S.** Global rise in human infectious disease outbreaks. *J. R. Soc. Interface* 11, e20140950, 2014
- Soucy SM, Huang J, Gogarten JP.** Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* 16, 472–82, 2015
- Stattner E, Vidot N.** *Social network analysis in epidemiology: Current trends and perspectives.* 5th International Research Challenges in Information Science (RCIS), Institute of Electrical and Electronics Engineers, Gosier, France, 2011

- Steel M.** A brief overview of phylogenetics and its terminology. *Trends Genet.* 21, 307–309, 2005
- Stevens JR, Schofield CJ.** Phylogenetics and sequence analysis--some problems for the unwary. *Trends Parasitol.* 19, 582-8, 2003
- Stoner BP, Whittington WL, Hughes JP, Aral SO, Holmes KK.** Comparative epidemiology of heterosexual gonococcal and chlamydial networks: implications for transmission patterns. *Sex. Transm. Dis.* 27, 215-23, 2000
- Streicker DG, Lemey P, Velasco-Villa A, Rupprecht CE.** Rates of viral evolution are linked to host geography in bat rabies. *PLoS Pathog.* 8, e1002720, 2012
- Strober MH.** Habits of the Mind: Challenges for Multidisciplinary Engagement. *Soc. Epistemol.* 20, 315-31, 2006
- Tamminen L-M, Söderlund R, Wilkinson DA, Torsein M, Eriksson E, Churakov M, Dicksved J, Keeling LJ, Emanuelson U.** Risk factors and dynamics of verotoxigenic *Escherichia coli* O157:H7 on cattle farms: An observational study combining information from questionnaires, spatial data and molecular analyses. *Prev. Vet. Med.* 170, 104726-36, 2019
- Tamura K, Nei M.** Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512–526, 1993
- Taylor LH, Latham SM, Woolhouse MEJ.** Risk factors for human disease emergence. *Philos. Trans. R. Soc. Lond. B* 356, 983-9, 2001
- Tedersoo L, Mett M, Ishida TA, Bahram M.** Phylogenetic relationships among host plants explain differences in fungal species richness and community composition in ectomycorrhizal symbiosis. *New Phytol.* 199, 822-31, 2013
- Thomas CM, Nielsen KM.** Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* 3, 711-21, 2005
- Thompson JN.** The evolution of species interactions. *Science* 284, 2116-8, 1999
- Thompson JN, Cunningham BM.** Geographic structure and dynamics of coevolutionary selection. *Nature* 417, 735-8, 2002

- Tian H, Cui Y, Dong L, Zhou S, Li X, Huang S, Yang R, Xu B.** Spatial, temporal and genetic dynamics of highly pathogenic avian influenza A (H5N1) virus in China. *BMC Infect. Dis.* 15, 54-66, 2015
- Tinsley M, Lewis FI, Brulisaue F.** Network modeling of BVD transmission. *Vet. Res.* 43, 11-16, 2012
- Valente TW.** Network Interventions. *Science* 337, 49-53, 2012
- VanderWaal KL, Atwill ER, Isbell LA, McCowan B.** Linking social and pathogen transmission networks using microbial genetics in giraffe (*Giraffa camelopardalis*). *J. Anim. Ecol.* 83, 406–414, 2014
- van Baalen M.** *Contact networks and the evolution of virulence — implications for virulence management.* In: The Adaptive Dynamics of Infectious Diseases: In Pursuit of Virulence Management. Cambridge University Press, Cambridge, UK, 2002
- van Belkum A, Struelens M, Visser A de, Verbrugh H, Tibayrenc M.** Role of Genomic Typing in Taxonomy, Evolutionary Genetics, and Microbial Epidemiology. *Clin. Microbiol. Rev.* 14, 547–560, 2001
- van Kerkhove MD, Vong S, Guitian J, Holl D, Mangtani P, San S, Ghani AC.** Poultry movement networks in Cambodia: implications for surveillance and control of highly pathogenic avian influenza (HPAI/H5N1). *Vaccine* 27, 6345-52, 2009
- van Segbroeck S, Santos FC, Pacheco JM.** Adaptive contact networks change effective disease infectiousness and dynamics. *PLoS Comput. Biol.* 6, 2010
- VanderWaal K, Enns EA, Picasso C, Packer C, Craft ME.** Evaluating empirical contact networks as potential transmission pathways for infectious diseases. *J. R. Soc. Interface* 13, 2016
- Volz E, Meyers LA.** Susceptible–infected–recovered epidemics in dynamic contact networks. *Proc. R. Soc. B Biol. Sci.* 274, 2925–2934, 2007
- Volz EM, Miller JC, Galvani A, Ancel Meyers L.** Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. *PLoS Comput. Biol.* 7, e1002042, 2011

- Voss JD, Mills DK, Myers JL, Remily ER, Richardson LL. Black band disease microbial community variation on corals in three regions of the wider Caribbean. *Microb. Ecol.* 54, 730-9, 2007
- Wasserman S, Faust K. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, UK, 1994
- Watts D, Strogatz S. Collective dynamics of ‘small-world’ networks. *Nature* 393, 440–442, 1998.
- Wehner J, Powell JR, Muller LAH, Caruso T, Veresoglou SD, Hempel S, Rillig MC. Determinants of root-associated fungal communities within Asteraceae in a semi-arid grassland. *J. Ecol.* 102, 425-36, 2014
- Weiss RA. Virulence and pathogenesis. *Trends Microbiol.* 10, 314–317, 2002
- Welch D, Bansal S, Hunter DR. Statistical inference to advance network models in epidemiology. *Epidemics* 3, 38-45, 2011
- Whelan S, Lio P, Goldman N. Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends Genet.* 17, 262-72, 2001
- World Health Organisation (WHO). *The World Health Report 2003: Shaping the Future. Chapter 5, SARS: Lessons from a new Disease*. World Health Organization, Geneva, Switzerland 2003
- World Health Organisation (WHO). *Global Health Estimates 2016: Disease burden by Cause, Age, Sex, by Country and by Region, 2000-2016*. Available at: https://www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html World Health Organisation, Geneva, Switzerland, 2018
- Woolhouse M, Donaldson A. Managing foot-and-mouth. *Nature* 410, 515-6, 2001
- Woolhouse ME, Shaw DJ, Matthews L, Liu WC, Mellor DJ, Thomas MR. Epidemiological implications of the contact network structure for cattle farms and the 20-80 rule. *Biol. Lett.* 1, 350-2, 2005
- Wylie JL, Jolly A. Patterns of chlamydia and gonorrhoea infection in sexual networks in Manitoba, Canada. *Sex. Transm. Dis.* 28, 14-24, 2001
- Xu RH, He JF, Evans MR, Peng GW, Field HE, Yu DW, Lee CK, Luo HM, Lin WS, Lin P, Li LH, Liang WJ, Lin JY, Schnur A. Epidemiologic clues to SARS origin in China. *Emerg. Infect. Dis.* 10, 1030-7, 2004

- Yang Z, Rannala B.** Molecular phylogenetics: principles and practice. *Nat. Rev. Genet.* 13, 303-14, 2012
- Yang T, Deng HW, Niu T.** Critical assessment of coalescent simulators in modeling recombination hotspots in genomic sequences. *BMC Bioinform.* 15, 3, 2014
- Young SG, Carrel M, Kitchen A, Malanson GP, Tamerius J, Ali M, Kayali G.** How's the Flu Getting Through? Landscape genetics suggests both humans and birds spread H5N1 in Egypt. *Infect. Genet. Evol.* 49, 293-9, 2017
- Ypma RJ, Bataille AM, Stegeman A, Koch G, Wallinga J, van Ballegooijen WM.** Unravelling transmission trees of infectious diseases by combining genetic and epidemiological data. *Proc. Biol. Sci.* 279, 444-50, 2012
- Yu ITS, Li Y, Wong TW, Tam W, Chan AT, Lee JHW, Leung DYC, Ho T.** Evidence of Airborne Transmission of the Severe Acute Respiratory Syndrome Virus. *N. Eng. J. Med.* 350, 1731–1739, 2004
- Zhang M, He H.** Parasite-mediated selection of major histocompatibility complex variability in wild brandt's voles (*Lasiopodomys brandtii*) from Inner Mongolia, China. *BMC Evol. Biol.* 13, 149, 2013
- Zharkikh A.** Estimation of evolutionary distances between nucleotide sequences. *J. Mol. Evol.* 39, 315–329, 1994
- Zuckerkandl E, Pauling L.** *Molecular disease, evolution, and genetic heterogeneity.* Academic Press, New York, 1962

Estimating the level of disease risk and biosecurity on commercial poultry farms in New Zealand

Sabrina S. Greening, Kerry Mulqueen, Thomas G. Rawdon, Nigel P. French, and M.
Carolyn Gates

3.1. Abstract

AIMS: To (i) collect and update baseline data on the contact patterns and biosecurity practices of farms within the New Zealand commercial poultry industry, (ii) investigate the relationship between the estimated farm-level contact risk and biosecurity practices, and (iii) identify important poultry health concerns that may be influencing the risk management behaviour of producers.

METHODS: A cross-sectional survey of all registered New Zealand commercial poultry operations was conducted in 2016 collecting information on farm demographics, biosecurity practices, and contact risk pathways. The quantitative survey responses were used to generate a subjective contact risk score based on the presence of eight potential disease transmission pathways and a subjective biosecurity score based on the frequency with which producers reported implementing seven common biosecurity measures. The correlation between the two scores was analysed with a Kruskal-Wallis rank sum test. Thematic analysis was performed on the qualitative free-text survey responses to further investigate producer opinions towards poultry health issues.

RESULTS: The survey response rate was 29.0% (120/414) from 57 (47.5%) broilers, 33 (27.5%) layers, 24 (20.0%) breeders, and 6 (5.0%) other poultry production types. The presence and absence of different contact risk pathways on each farm were highly variable both within and between each poultry sector. However, for both broiler and layer enterprises the greatest contribution to the contact risk score was associated with the movement of employees whereas, for breeder enterprises, the on- and off-farm movement of goods and services was considered the greater risk. Biosecurity adoption was generally low with only 14 (11.7%) reporting the use of all seven surveyed biosecurity measures. Overall, no significant correlation was found between the biosecurity score and contact risk score on each farm. Producer free-text responses showed a high level of

concern over the efficacy of biosecurity measures, highlighting an important future area of research.

CONCLUSIONS: The uptake of biosecurity measures in the New Zealand commercial poultry industry is highly variable but generally low despite the significant potential risk for diseases to spread through frequent between-farm contacts. This may be related to the low prevalence or absence of many important infectious poultry diseases in New Zealand leading farmers to believe there is a limited need to maintain good biosecurity as well as farmer uncertainty around the efficacy of different biosecurity measures. Further research is needed to understand barriers towards biosecurity adoption including evaluating the cost-effectiveness of biosecurity interventions.

KEYWORDS: Poultry health, Biosecurity, Risk ranking, Producer behaviour, Contact networks

3.2. Introduction

The emergence and spread of infectious diseases can have a devastating impact on commercial poultry industries, especially if the disease is zoonotic or foodborne in origin with potentially major implications on public health (Astill *et al.* 2018). To minimise the risk of disease introduction and spread, many different biosecurity measures can be implemented with some designed to target specific pathogens while others are more generalised. However, in many countries, it is widely recognised that a proportion of producers are failing to implement commonly recommended biosecurity practices (Moore *et al.* 2008; Kruger *et al.* 2009; Mankad, 2016) resulting in a huge variation in risk management practices between producers. Various factors are thought to play a role in a producer's decision to adopt different biosecurity measures, including physical farm characteristics (such as land area, flock size, number of neighbouring operations, company policies and the distance to the nearest road) (Lestari *et al.* 2012; Susilowati *et al.* 2013), producer characteristics (such as age, education level, gender, income, household size, years of experience, risk awareness and tolerance) (Racicot *et al.* 2012a; Akintunde and Adeoti, 2014), local factors (such as policy obligation, social expectations and local disease prevalence) (Funk *et al.* 2010; Itagaki, 2013; Hidano *et al.* 2018) and factors relating directly to implementation (such as the direct costs, public health benefits and the ease and practicality of adoption) (Fraser *et al.* 2010; Garforth *et al.* 2013). This makes it difficult to ensure a minimum level of biosecurity is maintained across all production premises and identify those farms that are of greatest risk in the event of a disease outbreak.

Being able to classify production premises based on their risk for disease introduction and spread is an important step in both the development of risk-based control and surveillance strategies as well as in many disease transmission models (Niemi *et al.* 2009; Van Steenwinkel *et al.* 2011). Despite this, there is only a limited number of studies that have tried to quantitatively assess the level of risk on production premises with

consideration for more than just one infectious agent (Bridges *et al.* 2007; Lewerin *et al.* 2015). However, more recently, a growing number of studies have used network-based approaches to identify high-risk sites that would be important to target in the event of a disease outbreak (Martínez-López *et al.* 2009; Rautureau *et al.* 2012; Sánchez-Matamoros *et al.* 2013). Nevertheless, only a limited number take into account on-farm biosecurity practices despite their clear influence on the risk of disease introduction and spread. This may be due to the difficulties in accurately assessing the level of biosecurity across different production premises (Nespeca *et al.* 1997; East, 2007) as well as a general unawareness as to how effective individual biosecurity measures are at preventing or reducing the spread of an infectious agent.

New Zealand's poultry industry offers a unique opportunity to study the relationship between disease risk and biosecurity adoption mainly due the industry's comparatively small size with approximately 119 million meat chickens (*i.e.*, broilers) raised annually, 3.5 million laying hens excluding a further 3 million replacements raised each year, and 2.5 million meat and layer breeder birds (Anonymous, 2018). In addition, there are previous studies that have focused on characterising the network structure within the New Zealand commercial poultry industry (Lockhart *et al.* 2010) that offers a great opportunity for comparison, with further studies that have quantified the frequency of routine biosecurity practices within the commercial poultry industry and investigated the importance of different disease risk pathways (Rawdon *et al.* 2007; Rawdon *et al.* 2008). A comparison with these studies would be both timely and relevant given the recent emergence and rapid spread of a previously unidentified antimicrobial resistant strain of *Campylobacter jejuni* that has since been responsible for both sporadic and outbreak-associated human cases of campylobacteriosis (Zhang *et al.* 2010; Taveirne *et al.* 2017). This emerging strain was found across all of the major poultry suppliers responsible for serving over 90% of the industry (Muellner *et al.* 2016). This represented a major epidemiological shift as previous *Campylobacter* strains have been strongly associated

with individual poultry suppliers (Müllner *et al.* 2010); highlighting a need for updated information on biosecurity practices and contact patterns.

As part of efforts to better understand the mechanisms responsible for this epidemiological shift a cross-sectional survey was conducted across all commercial poultry producers in New Zealand in order to (i) collect and update baseline data on farm biosecurity practices and contact risk pathways, (ii) investigate the relationship between contact risk pathways and on-farm biosecurity, and (iii) identify important poultry health concerns for producers that may have contributed to a shift in biosecurity practices.

3.3. Materials and Methods

3.3.1. Survey implementation

Contact details for commercial poultry enterprises in New Zealand were obtained from the Poultry Industry Association of New Zealand (PIANZ) and the Egg Producers' Federation (EPF). The database was accessed on June 2016 and listed 426 enterprises including hatcheries, breeding and rearing units, layer farms and broiler production units. This database was believed to capture the majority of commercial poultry enterprises since it is a mandatory requirement for layer farms to be members of EPF under the Commodity Levies (Eggs) Order 2009, whilst PIANZ membership represents over 99% of the country's chicken meat producers. Enterprises no longer in production or with no production facilities (*i.e.*, head offices) were removed along with duplicate records leaving 414 records believed to be active poultry producers. A pre-survey sensitisation e-mail was distributed through the PIANZ and EPF e-mail lists prior to paper copies of the questionnaire being mailed to the 414 active poultry producers identified including all broiler, layer, turkey, and duck enterprises. Paper copies were mailed on 15th June 2016 and the survey remained open until 15th December 2016 with two reminder e-mails sent to non-respondents during this period. The study was judged

to be low risk thorough peer evaluation and consequently was not formally reviewed by any of the University's Human Ethics Committees.

3.3.2. Survey design

The 11-page survey was based on a previous questionnaire administered to New Zealand poultry producers by Lockhart and colleagues (2010) in 2006 and modified in collaboration with PIANZ, EPF and the Ministry for Primary Industries (MPI) (Lockhart *et al.* 2010). The survey was designed to collect information on the farm demographics, contact patterns, and biosecurity practices of New Zealand commercial poultry operations and a copy of the complete survey questionnaire can be found in Appendix A. To summarise, the farm demographic variables included in the survey questionnaire aimed at capturing the (i) name and mailing address of the farm, (ii) spatial location of the farm, (iii) types of poultry produced, (iv) average number of birds present on the farm, (v) total number of poultry sheds on the farm, (vi) average maximum capacity of each shed, (vii) total farm capacity, (viii) predominant housing type, (ix) flow management of birds, (x) length of an average production cycle, (xi) downtime between production cycles, and (xii) number of full-time and part-time workers on the farm.

The contact variables aimed at capturing all movements of (i) transporting vehicles, (ii) feed, (iii) live birds and hatching eggs, (iv) table eggs and poultry products, (v) personnel, and (vi) poultry waste and litter, on- and off-farm over the previous one-year period. For each type of movement, producers were asked to provide additional details regarding the name and business location of each source or destination company, the direction of the movement (*i.e.*, onto or off-farm), the type and quantity of any products transferred by the movement, and the frequency of movements. It was also recorded whether the farms had direct contact with backyard poultry flocks through the sale of poultry, including end-of-lay or point-of-lay birds, or adjacent backyard poultry flock.

The biosecurity variables aimed at capturing the (i) allocation of workers across the farm, (ii) biosecurity measures implemented on farm, (iii) sharing and cleaning of equipment and vehicles between farms, (iv) sources of water supply and water treatment, and (v) presence of wild birds and waterfowls either in the same areas as poultry or on farm ponds and waterways. Lastly, all producers were asked to indicate their views on the importance of different poultry health issues using various Likert scales and open-ended questions.

3.3.3. Data processing

All responses were entered into a Microsoft Access database by two separate individuals and cross-checked for discrepancies. Ambiguous answers (*i.e.*, indicating both “yes” and “no” for a given question) were recorded as a missing response before data were imported into the R statistical software (R-Development Core Team, 2010) for processing. Firstly, enterprises were categorised in accordance with their primary production type resulting in four categories: broiler, layer, breeder, and all other poultry. If survey respondents indicated that multiple production types were present within a single enterprise, the primary production type was considered to be the one where over 80% of the poultry on-site were contributing to this purpose. The ‘other’ category combined pullet, duck, and turkey enterprises since there were relatively few of these in the study sample. A postal address for each of the study farms was retrieved by using the addresses provided in the survey. These addresses were checked using Google Maps (2017), to make sure they corresponded to a poultry production unit (indicated by the presence of poultry sheds) and not the producer’s residential address. For non-responders, addresses were obtained from the PIANZ-EPF producer list and checked accordingly. All addresses were used to obtain coordinates that could be used to map the location of (i) all active poultry producers in the PIANZ-EPF database and (ii) the producers that responded to the survey.

3.3.4. Generating the farm-level contact risk score

A subjective-unweighted contact risk score was calculated for all farms that responded to the survey questionnaire with the aid of a risk ranking matrix (Table 3.1). This was used as a proxy for the farm's potential to acquire and/or spread infectious diseases through the contact network. Eight risk criteria were contained within the matrix each reflecting a potential contact pathway for disease transmission. Risk criteria were selected if they had been previously identified as a risk pathway in the study by Rawdon and colleagues (2007, 2008) or based on evidence in the current literature highlighting their role in a disease transmission pathway (Rawdon *et al.* 2007; Rawdon *et al.* 2008). For each surveyed poultry farm, the presence or absence of a contact was identified from survey responses by looking at the farm characteristics and management practices in regards to (i) the number and assignment of employees, (ii) the sharing and cleaning of equipment, (iii) the presences of wild birds and waterfowl, (iv) contact with non-commercial poultry, (v) litter management, and (v) the supply and treatment of drinking water.

In addition, five network graphs were constructed from the reported on- and off-farm movements relating to (i) feed, (ii) live birds and hatching eggs, (iii) poultry waste, litter, and dead birds, (iv) additional poultry products, (v) personnel and (vi) all on- and off-farm movements. Further details on how the network graphs were constructed have been provided in Appendix B; however, for the purpose of this study only the network graph constructed from all the on- and off-farm movements was used to perform a social network analysis (SNA) and generate a degree centrality score was for each node (*i.e.*, a surveyed poultry farm) with a full description of the complete SNA also provided in Appendix B. The degree centrality score was used as an additional risk criterion to try and capture the potential transmission risk due between farms due to the on- and off-farm movements of goods and services. One last risk criterion, the number of neighbouring commercial poultry farms, was also calculated without reference to survey responses. This criterion was selected to try and capture the potential transmission risk due to the

spatial proximity of neighbouring poultry producers, not including backyard poultry producers. For this study, two poultry enterprises were considered neighbours if they were within a 5km radius of each other. This was determined for each farm by drawing a circular window, with a 5km radius, around a centroid position marked by the farm coordinates that had been extracted from the PIANZ-EPF database. Poultry farms captured within this window were considered as a neighbouring poultry premises.

Table 3.1. The disease risk-ranking matrix showing the contribution of risk attributes within each of the eight-risk criterion to the estimated disease risk score.

Risk Criterion (#) Risk attribute	Likelihood of disease introduction and transmission			
	Minimal	Low	Moderate	High
Neighbouring farms	0	1-5	6-15	>15
Drinking-Water				
(1) Sources	(1) Water from low-risk sources ^c	(1) Water from low-risk sources ^c	(1) Water from high-risk sources ^c	(1) Water from high-risk sources ^c
(2) Treatment	(2) Water treated	(2) Water untreated	(2) Water treated	(2) Water untreated or only filtered
Litter management				
(1) Cleaning of poultry sheds	(1) Poultry sheds fully cleaned out and disinfected with replacement litter treated	(1) Poultry sheds fully cleaned out and disinfected, but replacement litter not treated	(1) Either poultry sheds fully cleaned but not disinfection, or sheds only partially cleaned and disinfected	(1) Sheds partially cleaned with no disinfection or litter treatment
(2) Litter moisture	(2) Litter just right	(2) Litter just right	(2) Litter too dry	(2) Litter too wet
(3) Downtime ^a	(3) >28 days	(3) 15-28 days	(3) 5-14 days	(3) <5 days

Table 3.1 continues next page

Table 3.1 continued

Risk Criterion (#) Risk attribute	Likelihood of disease introduction and transmission			
	Minimal	Low	Moderate	High
Non-commercial poultry				
(1) Number of bordering backyard poultry	(1) No bordering backyard farms	(1) 1 bordering backyard farm	(1) 2 bordering backyard farms	(1) ≥ 3 bordering backyard farms
(2) Sale of end-of-lay or point-of-lay birds	(2) No sales	(2) No sales	(2) Sale of end-of-lay or point-of-lay birds	(2) Sale of end-of-lay or point-of-lay birds
Wild birds and waterfowl				
(1) Birds present	(1) No birds present on the farm	(1) Birds present on farm ponds	(1) Birds present on farm ponds	(1) Birds present in production areas
(2) Distance of birds from the production area	(2) NA	(2) >50m from the production area	(2) ≤ 50 m from the production area	(2) Within the production area
Equipment				
(1) Shared	(1) No equipment shared	(1) Equipment shared	(1) Equipment shared	(1) Equipment shared
(2) Cleaned	(2) NA	(2) All equipment cleaned	(2) Only some equipment cleaned	(2) No equipment cleaned

Table 3.1 continues next page

Table 3.1 continued

Risk Criterion (#) Risk attribute	Likelihood of disease introduction and transmission			
	Minimal	Low	Moderate	High
Employees				
(1) Number	(1) 0	(1) 1-4	(1) 5-10	(1) >10
(2) Shed assignment	(2) NA	(2) Specific sheds	(2) Specific sheds	(2) >1 full farm
Degree centrality^b	0	1-10	11-25	>25

^a time in days between production cycles in which poultry sheds are left empty

^b sum of node in-degree and out-degree calculated from all movements going either on to or off farm

^c water collected from the town supply was considered low risk compared to water collected from roofs, rivers, streams, and springs which were considered high-risk sources

Each of the eight risk criteria were then ranked on a Likert-scale so that for each risk a node received a score of either zero (minimal risk), one (low risk), two (moderate risk) or three (high risk) (Table 3.1). The score was given based on a biological understanding of the poultry production system and the likelihood that the attribute would result in disease introduction or dissemination without consideration of a specific pathogen. Overall using this risk ranking, the total contact risk score could range between zero, indicating that the level of potential risk on a farm was minimal, to 24, the maximum level of potential risk.

3.3.5. Generating the farm-level biosecurity frequency score

A subjective biosecurity score was calculated for each surveyed farm from responses that indicated the frequency at which seven common biosecurity practices were implemented on farm. The seven surveyed biosecurity measures included the use of (i) dedicated coveralls for each shed, (ii) dedicated foot covers for each shed, (iii) footbaths at shed entrances, (iv) rodent bait stations, (v) bird-proofed housing, (vi) bird-proofed feed stores, and (vii) vehicle disinfection before entering the farm. These measures were

selected for inclusion in the survey on the basis that they are routine daily practices recommended as a minimum biosecurity requirement on broiler farms in the PIANZ biosecurity manual (PIANZ, 2015). For this study, all biosecurity measures were considered equally important due to the limited research comparing the effectiveness of biosecurity measures in mitigating disease transmission; making it difficult to give weights to individual biosecurity measures. In the survey, responders indicated the frequency of implementing each biosecurity measure on a five-point Likert-scale, making it easy to rank answers with no consideration of the efficiency of the control measure for preventing disease transmission. Responses were given a corresponding score ranging from zero (never), one (rarely), two (sometimes), three (often) and four (always). The sum of these indicators made up the biosecurity score for each farm, resulting in a range of zero (minimum level of biosecurity) to 28.0 (maximum level of biosecurity).

3.3.6. Statistical analysis

Basic descriptive statistics were conducted as appropriate using R statistical software (R-Development Core Team, 2010) to summarise survey results and make a comparison between the different poultry production types. To further investigate if there was a difference in the contact risk score and the biosecurity score between the different poultry production types, a Kruskal-Wallis rank sum test was performed. If this initial test was found to be significant ($p < 0.05$), an additional Dunn's (1964) test was then performed with multiple comparisons adjusted for using a Bonferroni adjustment. The relationship between the contact risk score and biosecurity score on each farm was visualised with five basic scatterplots plots showing (i) all survey respondents, (ii) only surveyed broiler enterprises, (iii) only surveyed layer enterprises, (iv) only surveyed breeder enterprises and (v) all other survey farms. The Pearson's correlation coefficient (PCC) was then calculated for each plot due to the non-normal distribution of the data. Lastly, to analyse the open-ended survey responses regarding producer opinions towards poultry health concerns and future research interests, word frequency queries were performed using

NVivo v12.0 (QSR International Pty *Ltd.*, 2018) in order to group responses. Groups were then examined manually to identify any common themes between producers.

3.4. Results

3.4.1. Poultry demographics

The postal survey was returned by 29.0% (120/414) of the active poultry producers registered in the PIANZ-EPF database as of December 2016. The breakdown of responses varied by production type with the two largest industry sectors, broilers and layers, unequally represented (Table 3.2). However, there was a representative geographical distribution of respondents from across New Zealand (Figure 3.1).

3.4.2. Contact patterns and contact risk scores

The contact risk pathways identified on each farm were highly variable both within and between each of the poultry sectors resulting in a wide range of contact risk scores (Figure 3.2). For example, when considering the potential risk from contact with non-commercial backyard poultry, 84.2% (101/120) of the study farms had minimal risk (Table 3.3). Out of the remaining farms, 94.7% (18/19) were layer enterprises whose risk with non-commercial backyard poultry was largely due to the sale of point-of-lay and end-of-lay birds. Despite this variation, the greatest risk for both layer and broiler enterprises was from the potential movement of employees between sheds, with 78.8% (26/33) of layer operations and 98.2% (56/57) of broiler operations indicating that they have a number of employees assigned to work across the whole farm (Table 3.3). For breeder enterprises, the greatest risk was attributed to the on- and off-farm movement of goods and services with 95.8% (23/24) having a network degree centrality measure greater than 25 (Appendix B, Figure B1) while the number of neighbouring farms, litter management practice and water sources were only considered to present a moderate level of risk across all the study farms.

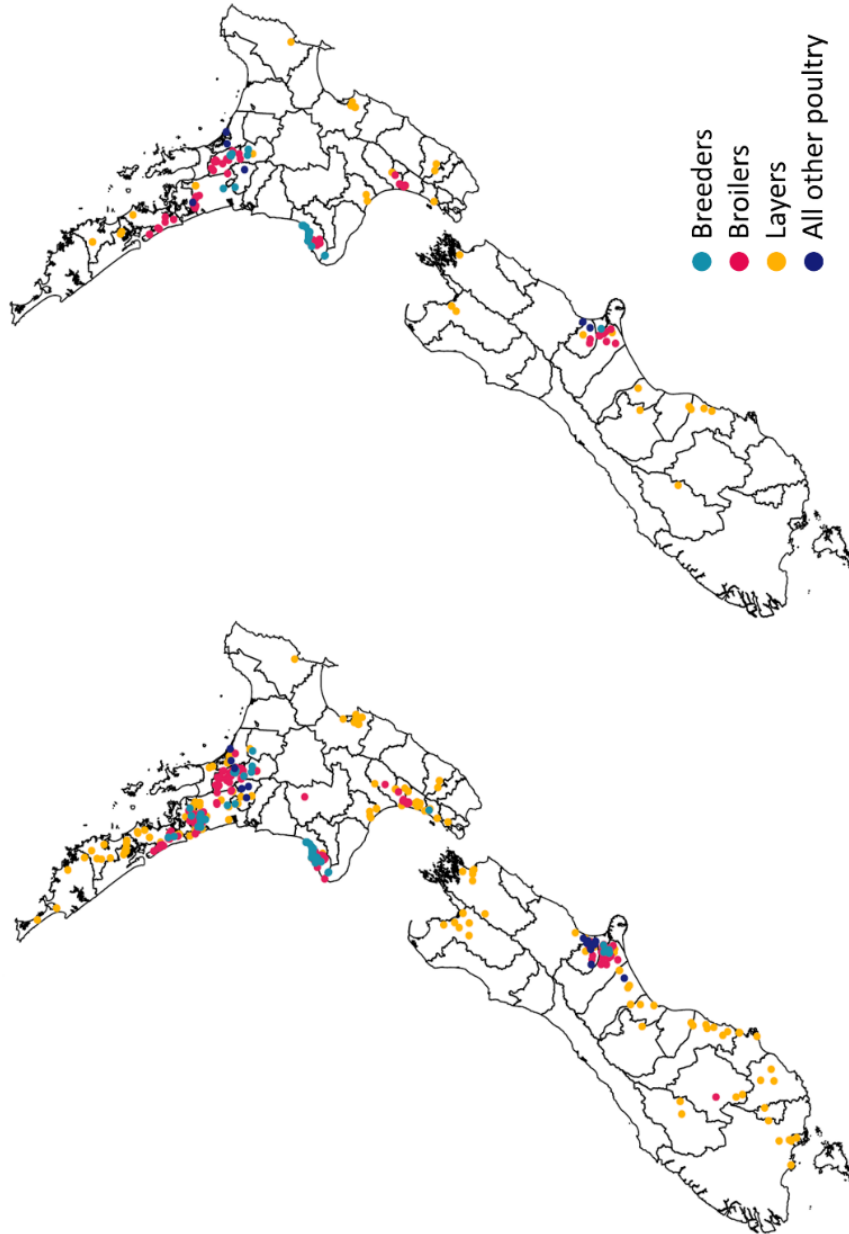


Figure 3.1. (a) The distribution of the 414 active poultry producers register with either the Poultry Industry Association of New Zealand (PIANZ) or the Egg Producers' Federation (EPF) as of June 2016, and (b) the subset of poultry producers who responded to the 2016 poultry survey including 24 breeder enterprises (light blue), 57 broiler enterprises (pink), 33 layer enterprises (yellow), and 6 other poultry enterprise (navy).

Table 3.2. Summary of demographic information amongst 120 producers in the New Zealand commercial poultry industry all of whom responded to the 2016 poultry survey.

	Poultry production type					
	Broilers	Layers ^d	Breeders	Pullets	Ducks	Turkeys
Number of survey responders (%)	57 (47.5)	33 (27.5)	24 (20.0)	2 (1.7)	3 (2.5)	1 (0.8)
Sector response (%)^a	57/157 (36.3)	33/169 (19.5)	24/55 (43.6)	2/15 (13.3)	3/6 (50.0)	1/11 (9.1)
Housing (%)						
<i>Free-range</i>	10/57 (17.5)	22/33 (66.7)	1/24 (4.2)	1/2 (50.0)	0/3 (0.0)	1/1 (100)
<i>Colony</i>	NA ^c	4/33 (12.1)	0/24 (0.0)	0/2 (0.0)	0/3 (0.0)	0/1 (0.0)
<i>Barn</i>	43/57 (75.4)	2/33 (6.1)	23/24 (95.8)	1/2 (50.0)	3/3 (100)	0/1 (0.0)
<i>Mixed^b</i>	1/57 (1.8)	5/33 (15.2)	0/24 (0.0)	0/2 (0.0)	0/3 (0.0)	0/1 (0.0)
Median number of poultry per farm (range)	97820 (19500-1000000)	8750 (20-150000)	15000 (5600-140000)	2500 (50-35000)	14000 (1000-18000)	22000 (-)
Median number of sheds per farm (range)	4.0 (2-12)	3.0 (1-14)	2.0 (1-24)	1.0 (1-5)	2.5 (2-3)	4.0 (-)

Table 3.2 continues next page

Table 3.2 continued

	Poultry production type					
	Broilers	Layers ^d	Breeders	Pullets	Ducks	Turkeys
Flow of birds (%)						
<i>All-in-all-out</i>	40/57 (70.2)	11/33 (33.3)	22/24 (91.7)	2/2 (100)	3/3 (100)	0/0 (0.0)
<i>Multiple ages</i>	16/57 (28.1)	15/33 (45.5)	2/24 (8.3)	0/0 (0.0)	0/0 (0.0)	0/0 (0.0)
<i>Mixed</i>	1/57 (1.8)	7/33 (21.2)	0/24 (0.0)	0/0 (0.0)	0/0 (0.0)	1/1 (100)
Median days of production cycle (range)	42.0 (18-56)	420 (42-630)	294 (140-560)	180 (-)	70.5 (43-98)	40.0 (-)
Median days downtime length (range)	8.0 (0-18)	14.0 (0-120)	42.0 (35-360)	26.5 (18-35)	8.5 (7-10)	9.0 (-)
Median part-time employees per farm	1.0	3.0	1.0	0.5	1.0	1.0
Median full-time employees per farm	1.0	1.0	4.0	3.0	1.0	1.0

^a Sector includes the 414 active poultry farms identified in the PIANZ-EPF database as of October 2016

^b Colony housing not applicable to broilers as all broiler/meat chickens must be on the floor in sheds or free-range. Despite this 3/57 (5.3%) of broiler producers reported colony housing, therefore, for the purpose of this study, these responses were considered as mistakes by the producer and have not been included

^c Mixed production systems include farms that indicated having both free-range birds and either birds in barns or colony housing

^d Includes mixed pullet and layer enterprises

- A range cannot be calculated as only one response has been given

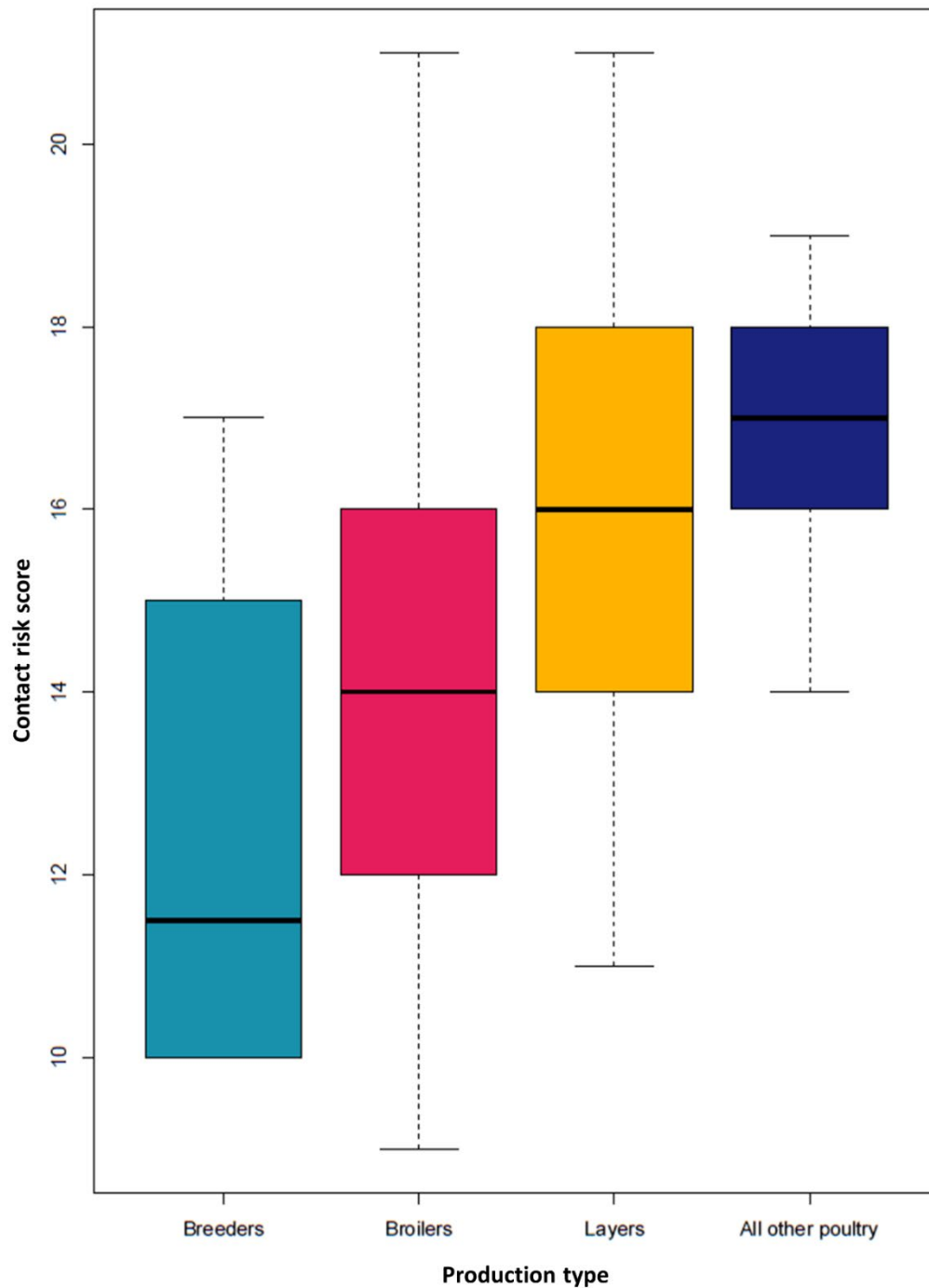


Figure 3.2. Boxplot showing the estimated risk score for 120 poultry producers in New Zealand's commercial poultry industry including 33 layer enterprises, 57 broiler enterprises, 24 breeder enterprises and 6 enterprises representing either duck, turkey or pullet operations. The values in the graph represent the minimum, maximum, median, first quartile and third quartile in the data with the median value (\tilde{x}) printed below each plot

Table 3.3. Contribution of risk criteria to the estimated disease risk score amongst 120 producers in the New Zealand commercial poultry industry including 33 layer enterprises (including mixed pullet and layer operations), 57 broiler enterprises, 24 breeder enterprises and 6 other poultry enterprises (including duck, turkey and pullet operations). A risk score of zero indicates a minimal risk of disease introduction or spread where a score of 3 indicates a maximum level risk.

Risk Criterion	Risk Score	Total number of farms (% within sector)			
		Layers	Broilers	Breeder	All other poultry
Number of neighbouring farms	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	1	29 (87.9)	32 (56.1)	14 (58.3)	5 (83.3)
	2	2 (6.1)	15 (26.3)	10 (41.7)	1 (16.7)
	3	2 (6.1)	10 (17.5)	0 (0.0)	0 (0.0)
Water source and treatment	0	8 (24.2)	50 (87.7)	20 (83.3)	1 (16.7)
	1	14 (42.4)	3 (5.3)	4 (16.7)	5 (83.3)
	2	6 (18.2)	4 (7)	0 (0.0)	0 (0.0)
	3	5 (15.2)	0 (0.0)	0 (0.0)	0 (0.0)
Litter management	0	1 (3.0)	0 (0.0)	0 (0.0)	0 (0.0)
	1	16 (48.5)	38 (66.7)	21 (87.5)	3 (50)
	2	9 (27.3)	19 (33.3)	2 (8.3)	1 (16.7)
	3	7 (21.2)	0 (0.0)	1 (4.2)	2 (33.3)
Contact with backyard poultry	0	15 (45.5)	57 (100)	24 (100)	5 (83.3)
	1	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	2	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	3	18 (54.5)	0 (0.0)	0 (0.0)	1 (16.7)
Contact with wild birds and waterfowl	0	11 (33.3)	26 (45.6)	17 (70.8)	1 (16.7)
	1	1 (3.0)	6 (10.5)	2 (8.3)	1 (16.7)
	2	0 (0.0)	4 (7.0)	4 (16.7)	1 (16.7)
	3	21 (63.6)	21 (36.8)	1 (4.2)	3 (50.0)
Equipment sharing	0	24 (72.7)	20 (35.1)	2 (8.3)	3 (50.0)
	1	4 (12.1)	20 (35.1)	21 (87.5)	1 (16.7)
	2	0 (0.0)	3 (5.3)	1 (4.2)	0 (0.0)
	3	5 (15.2)	14 (24.6)	0 (0.0)	2 (33.3)

Table 3.3 continues next page

Table 3.3 continued

Risk Criterion	Risk Score	Total number of farms (% within sector)			
		Layers	Broilers	Breeder	All other poultry
Number of employees	0	2 (6.1)	3 (5.3)	0 (0.0)	0 (0.0)
	1	15 (45.5)	48 (84.2)	5 (20.8)	5 (83.3)
	2	10 (30.3)	6 (10.5)	12 (50.0)	1 (16.7)
	3	6 (18.2)	0 (0.0)	7 (29.2)	0 (0.0)
Network degree centrality	0	1 (3.0)	0 (0.0)	0 (0.0)	0 (0.0)
	1	18 (54.5)	5 (8.8)	0 (0.0)	1 (16.7)
	2	9 (27.3)	16 (28.1)	1 (4.2)	0 (0.0)
	3	5 (15.2)	36 (63.2)	23 (95.8)	5 (83.3)

The complete breakdown of the different contact risk pathways reported in the survey can be found in Appendix B (Tables B4-B8) in addition to summaries on each of the reconstructed networks graphs (Appendix B, Tables B9) and statistics from the SNA (Appendix B, Tables B10). Overall, the mean contact risk score was 14.6 (median: 15, range: 9-21) across all farms. The contribution of different risk criteria to the final score is presented in Table 3.3. Results from the Dunn's test (Appendix B, Table B3) show that there was a significant difference between the production types ($p < 0.0001$) with duck, pullet and turkey enterprises having the greatest average risk score (mean: 16.7, median: 17, range: 14-19) followed by layer enterprises (mean: 15.9, median: 16, range: 11-21), broiler enterprises (mean: 14.2, median: 14, range: 9-21), and finally breeder enterprises (mean: 12.4, median: 11, range: 10-17).

3.4.3. Biosecurity practices and biosecurity frequency scores

The reported use of each biosecurity measure varied greatly (Figure 3.3). For example, 95.8% (115/120) of the survey respondents indicated 'always' using rodent bait stations in comparison to only 22.5% (27/120) that indicated 'always' disinfecting vehicles. In addition to rodent bait stations, the use of bird-proofed housing and bird-proofed feed stores were reportedly used by over 90% of all the survey respondents whilst disinfecting vehicles, dedicated shed overalls and footbath were the least reported (Figure 3.3) with

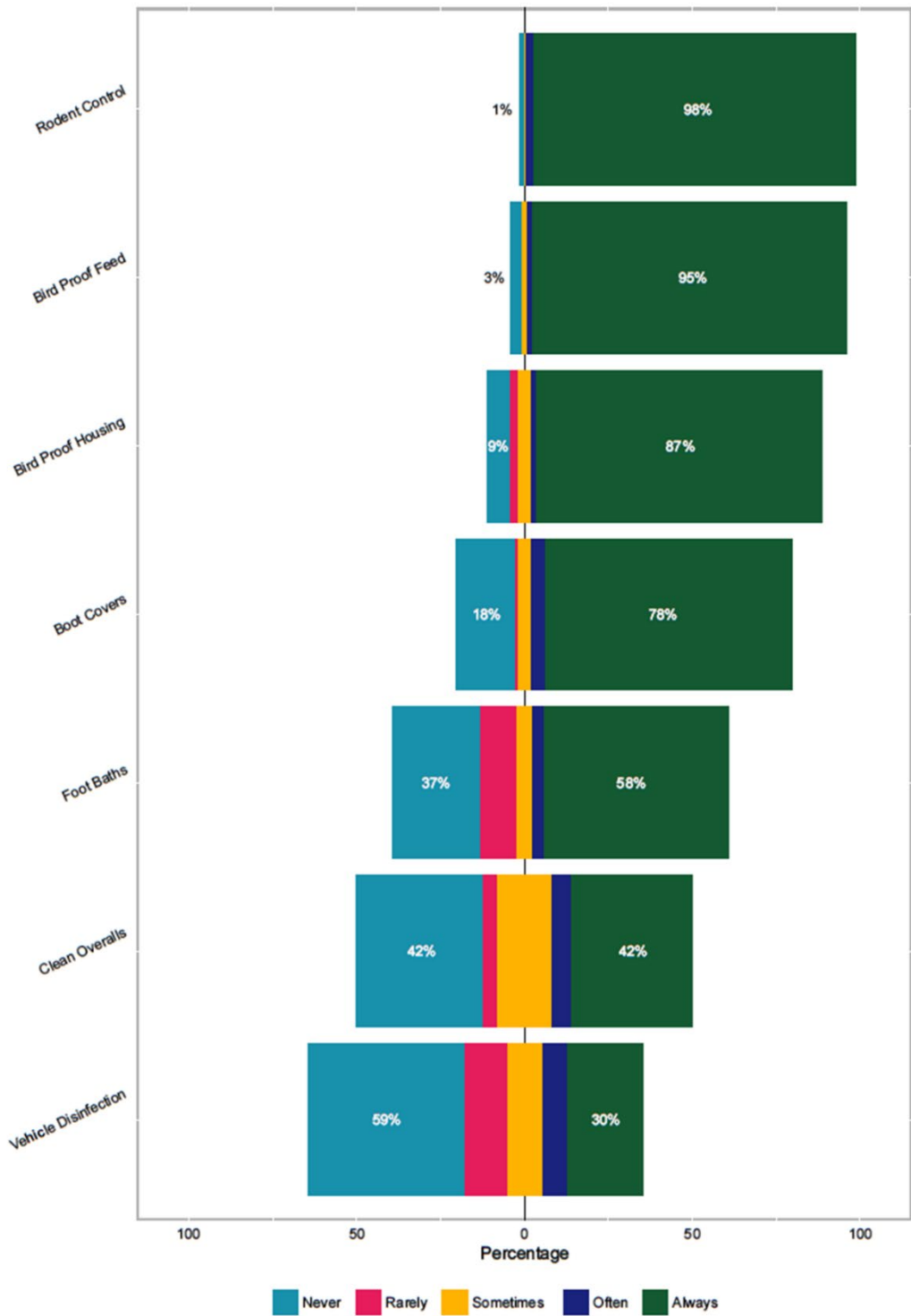


Figure 3.3. The reported frequency of implementing seven common biosecurity measures amongst 120 poultry producers in New Zealand's commercial poultry industry.

responses varying between the production types (Appendix B, Table B1). Overall, the mean biosecurity score was 20.4 (median: 22, range: 0-28) with the largest heterogeneity seen within the layer sector (Figure 3.4). Results from the Dunn's test (Appendix B, Table B2) show that there was a significant difference in the biosecurity scores between the production types ($p < 0.0001$) with breeder enterprises having the greatest average biosecurity score (mean: 24.3, median: 22, range: 20-28) followed by broiler enterprises (mean: 22.4, median: 23, range: 16-28), duck, pullet and turkey enterprises (mean: 20.8, median: 20, range: 19-23), and finally layer enterprises (mean: 14.0, median: 14, range: 0-26).

3.4.4. Association between the contact risk and biosecurity scores

The relationship between the estimated contact risk score and the biosecurity score (Figure 3.5) resulted in a PCC of -0.23 ($p = 0.01$, $df = 118$, 95% CI = -0.39 to -0.05) indicating there is no correlation between the number of contact risk pathways and the level of on-farm biosecurity, although the strength of this relationship varies when considering each production type individually (Appendix B, Figure B2).

3.4.5. Producer concerns and opinions about poultry health

When asked about six common poultry health concerns, the majority of survey respondents indicated being either 'not at all' concerned or 'extremely' concerned with very few responses in-between these extremes (Table 3.4). However, the proportion of responses in each of the two categories differed between the different production types with a higher proportion of layer enterprises showing little to no concern for campylobacter (42.4%, 14/33), salmonella (36.4%, 12/33), avian influenza (36.4%, 12/33), coccidiosis (54.5%, 18/88) or antimicrobial resistance (63.6%, 21/33) in comparison to the majority of broiler enterprises who showed the least concern for welfare (19.3%, 11/57) and antimicrobial resistance (15.8%, 9/57).

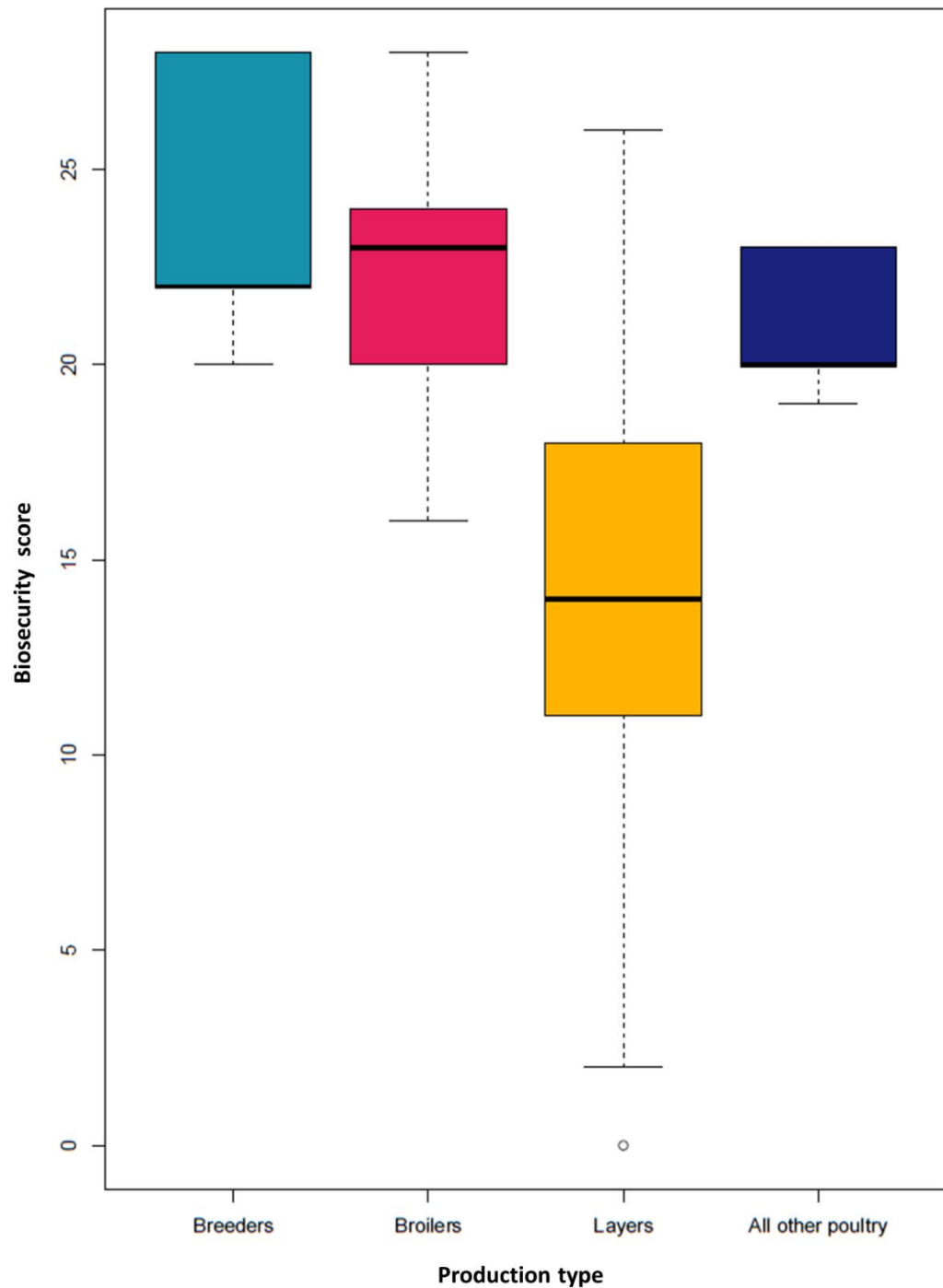


Figure 3.4. Boxplot showing the estimated biosecurity score for 120 poultry producers in New Zealand's commercial poultry industry including 33 layer enterprises, 57 broiler enterprises, 24 breeder enterprises and 6 enterprises representing either duck, turkey or pullet operations. The values in the graph represent the minimum, maximum, median, first quartile and third quartile in the data with the median value (\tilde{x}) printed below each plot.

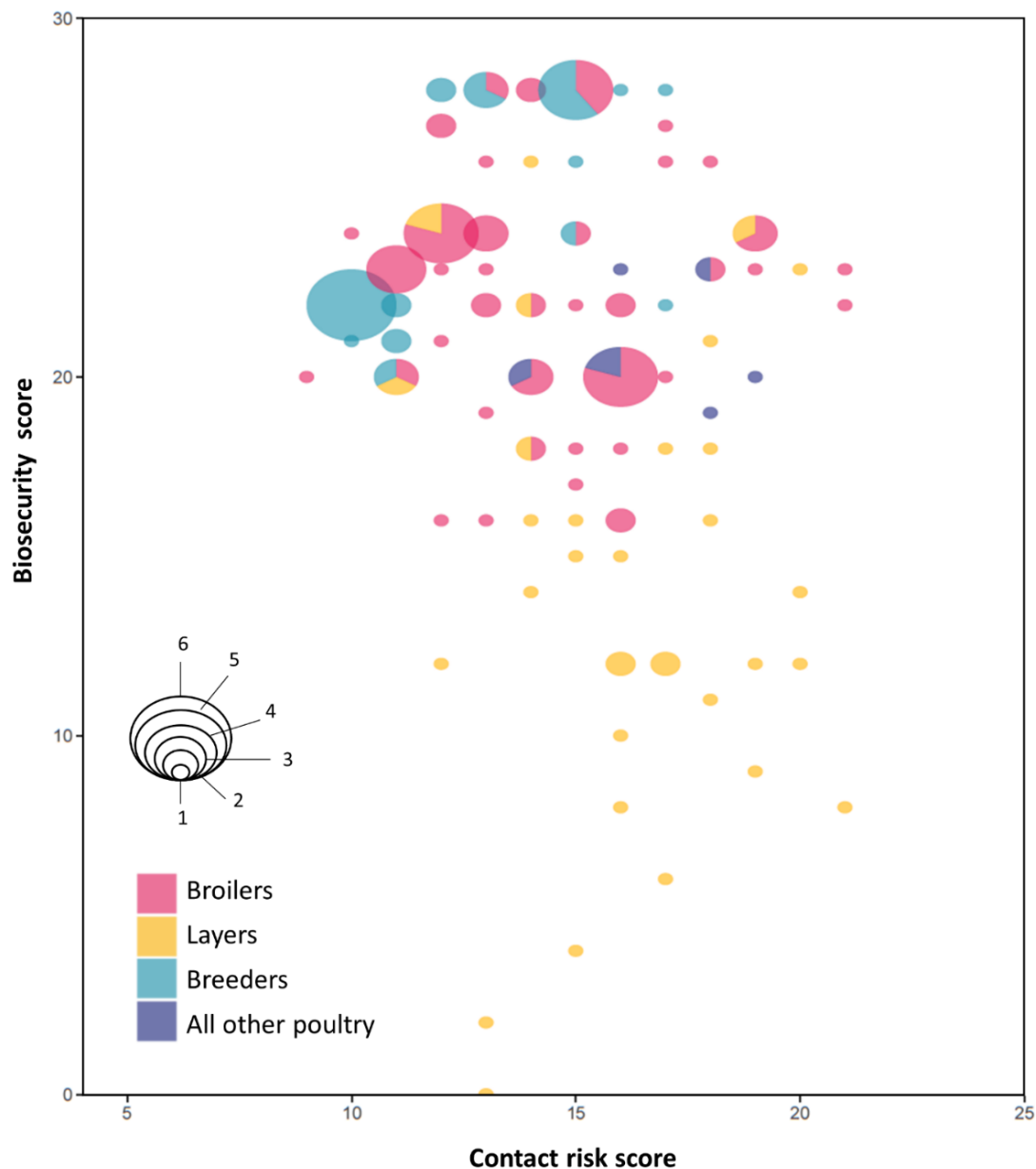


Figure 3.5. Scatter plot showing the relationship between the estimated biosecurity score, calculated from the reported frequency of implementing seven common biosecurity practices, and the estimated disease risk score for 120 poultry producers in New Zealand's commercial poultry industry including 33 layer enterprises, 57 broiler enterprises, 24 breeder enterprises and 6 enterprises representing either duck, turkey or pullet operations. Point circumference is proportional to farm frequency whilst the pie chart shows the breakdown of farms by production type. Pearson's correlation coefficient (PCC) has been calculated and 95% confidence intervals are shown for a p -value <0.01 .

Table 3.4. Level of concern over poultry health issues amongst 120 producers in the New Zealand commercial poultry industry including 33 layer enterprises (including mixed pullet and layer operations), 57 broiler enterprises, 24 breeder enterprises and 6 other poultry enterprises (including duck, turkey or pullet operations).

		Number of Farms (%)			
		Layers	Broilers	Breeder	All other poultry
Campylobacter	<i>Not at all</i>	14 (42.4)	6 (10.5)	12 (50.0)	2 (33.3)
	<i>Slightly</i>	7 (21.2)	8 (14.0)	11 (45.8)	0 (0.0)
	<i>Somewhat</i>	3 (9.1)	7 (12.3)	0 (0.0)	1 (16.7)
	<i>Moderately</i>	3 (9.1)	16 (28.1)	0 (0.0)	1 (16.7)
	<i>Extremely</i>	6 (18.2)	20 (35.1)	1 (4.2)	2 (33.3)
Salmonella	<i>Not at all</i>	12 (36.4)	7 (12.3)	11 (45.8)	2 (33.3)
	<i>Slightly</i>	6 (18.2)	6 (10.5)	1 (4.2)	0 (0.0)
	<i>Somewhat</i>	5 (15.2)	8 (14.0)	0 (0.0)	1 (16.7)
	<i>Moderately</i>	2 (6.1)	16 (28.1)	0 (0.0)	1 (16.7)
	<i>Extremely</i>	8 (24.2)	20 (35.1)	12 (50.0)	2 (33.3)
Avian Influenza	<i>Not at all</i>	12 (36.4)	7 (12.3)	12 (50.0)	2 (33.3)
	<i>Slightly</i>	7 (21.2)	9 (15.8)	0 (0.0)	1 (16.7)
	<i>Somewhat</i>	2 (6.1)	7 (12.3)	0 (0.0)	0 (0.0)
	<i>Moderately</i>	4 (12.1)	6 (10.5)	1 (4.2)	1 (16.7)
	<i>Extremely</i>	8 (24.2)	28 (49.1)	11 (45.8)	2 (33.3)
Coccidiosis	<i>Not at all</i>	18 (54.5)	4 (7.0)	13 (54.2)	2 (33.3)
	<i>Slightly</i>	5 (15.2)	8 (14.0)	0 (0.0)	2 (33.3)
	<i>Somewhat</i>	5 (15.2)	11 (19.3)	0 (0.0)	0 (0.0)
	<i>Moderately</i>	1 (3.0)	15 (26.3)	11 (45.8)	1 (16.7)
	<i>Extremely</i>	4 (12.1)	19 (33.3)	0 (0.0)	1 (16.7)
Antimicrobial resistance	<i>Not at all</i>	21 (63.6)	9 (15.8)	12 (50.0)	2 (33.3)
	<i>Slightly</i>	5 (15.2)	6 (10.5)	1 (4.2)	2 (33.3)
	<i>Somewhat</i>	2 (6.1)	13 (22.8)	11 (45.8)	0 (0.0)
	<i>Moderately</i>	1 (3.0)	13 (22.8)	0 (0.0)	1 (16.7)
	<i>Extremely</i>	4 (12.1)	16 (28.1)	0 (0.0)	1 (16.7)
Welfare	<i>Not at all</i>	16 (48.5)	11 (19.3)	11 (45.8)	2 (33.3)
	<i>Slightly</i>	2 (6.1)	4 (7.0)	0 (0.0)	0 (0.0)
	<i>Somewhat</i>	1 (3.0)	1 (1.8)	0 (0.0)	0 (0.0)
	<i>Moderately</i>	4 (12.1)	8 (14.0)	0 (0.0)	2 (33.3)
	<i>Extremely</i>	10 (30.3)	33 (57.9)	13 (54.2)	2 (33.3)

Seventy producers also provided free-text responses when asked about the biggest concerns facing their operations with several major themes emerging. First, producers were concerned that the new animal welfare legislation was based on public opinion rather than science, and worried about the impact of the changes on farm production levels as well as biosecurity; particularly when switching to free-range housing systems where their poultry would have greater contact with wildlife. Second, several producers were also concerned that urbanisation in their region and increased foot traffic from the general public was leading to increased contact with backyard poultry operations. This was seen as a potential pathway for both endemic diseases to spread between farms but also exotic diseases to enter the New Zealand commercial poultry industry if there were to be a lapse in border biosecurity with foreign visitors frequently utilising public walkways. Third, producers recognized that the movements of personnel and vehicles were a major risk for disease introductions and expressed concerns over effectively managing biosecurity. Lastly, a few producers also cited general issues with labour management and staffing.

Forty-five producers further provided free-text responses when asked what areas of future research would be most beneficial to their farms. There were many responses concerning improved strategies for managing litter on farm, developing a better understanding into which biosecurity practices (including vaccination, disinfection, and border control) are most effective in preventing endemic diseases from spreading between farms as well as controlling exotic disease from crossing the border and generating more evidence around the strengths and weaknesses of different housing systems. Three producers cited a need for learning how to better manage public perceptions about poultry production and welfare whilst two producers also wanted more research around how best to prevent foreign diseases from being introduced into New Zealand.

3.5. Discussion

The study results suggest that there are a significant number of contacts between commercial poultry farms in New Zealand, which have the potential to act as disease transmission pathways. However, despite these risks, a large proportion of the farms also had a relatively low uptake of biosecurity measures aimed at preventing disease transmission from occurring through these contacts. Overall, no association was found between the estimated level of risk and the estimated level of on-farm biosecurity for any poultry sector despite notable differences in both the relative importance of different contact risk pathways and biosecurity adoption rates between each sector with the contact of wild birds and backyard poultry contributing largely to the risk score for the majority of layer enterprises in comparison to both breeder and broiler enterprises whose greatest risk was through the movement of transporting vehicles and employees. Layer enterprises also had the lowest average biosecurity score with many indicating ‘never’ having used protective overalls, boot covers or footbaths, all of which, were reported frequently by broiler and breeder enterprises. These major differences between the poultry sectors are likely related to the differences in the industry demographic structures. For example, for breeder enterprises, a large proportion of the contact risk score was associated with the on- and off-farm movement of goods and services. This large number of movements is most likely due to the transfer of live birds and hatching eggs since the New Zealand poultry industry is highly vertically integrated and relies on only a small number of breeding farms and hatcheries to service the majority of broiler farms.

One of the most frequent contact risk pathways across all of the surveyed farms was the movement of employees. This result is similar to that reported by Rawdon and colleagues (2007, 2008) who also identified a large number of between-farms contacts was due to the frequent movement of personnel (Rawdon *et al.* 2007; Rawdon *et al.* 2008). However, there were other minor differences in the previously reported contact risk pathways compared with our present study. For example, untreated drinking water from high-risk

sources, such as rainwater, rivers, streams, and springs, was previously identified to be a major risk pathway for the entry of waterborne exotic pathogens. In our study, there was a slight increase in the proportion of farms that treated their drinking water, particularly in the layer industry. However, since most of the farms in our study sourced water from low-risk sources, most commonly bores, the probability of pathogen introduction via this pathway is very low even with a large proportion of farms leaving water untreated. Another important risk factor identified in the previous study was the sharing of equipment, whereas the majority of farms in our study reported that they did not share equipment with other farms and, of those that did, a large proportion indicated that they cleaned the equipment after its return to minimise the potential risk of disease introduction and spread. In comparison, the patterns of biosecurity adoption between the two studies has remained fairly constant, including some of the observed differences between the poultry sectors. For example, in both studies, a greater proportion of broiler enterprises reported the use of footbaths and protective clothing compared with layer enterprises whereas the use of rodent bait stations, bird-proofed housing and bird-proofed feed stores continued to be implemented by the majority of surveyed farms in both sectors (Rawdon *et al.* 2007; Rawdon *et al.* 2008).

In our current study, these differences in biosecurity practices have resulted in a greater biosecurity score among broiler and breeder enterprises in comparison to both layer enterprises and all ‘other’ poultry producers; suggesting that broiler and breeder enterprises may be more proactive in adopting biosecurity measures. This higher adoption rate may be for many reasons for instance, breeder enterprises may have pressure to maintain a high level of biosecurity given they supply many downstream farms particularly as the industry tries to reduce the use of antimicrobials meaning good biosecurity practices on breeder enterprises and hatcheries have become even more important (Anonymous, 2017). For broiler producers a large concern will be those pathogens responsible for foodborne illnesses, such as *Campylobacter*, with New Zealand

having one of the highest rates of human campylobacteriosis in comparison to other industrialised countries (Olson *et al.* 2008). In comparison, the primary concern for the layer industry will most likely be *Salmonella*, which has a comparatively much lower prevalence overall, reducing the perceived risk of disease; a factor known to influence biosecurity adoption rates (Hidano *et al.* 2018). The higher average biosecurity score among broilers could also be an artefact of the survey design, for example, many breeder enterprises are known to have on-farm showers to reduce the risk of disease introduction via the movement of employees, however, this practice was not included in the pre-determined list of biosecurity procedures in the survey-questionnaire. Instead only a limited number of biosecurity measures were included with no space provided for producers to expand on unidentified biosecurity procedures.

For our study, the biosecurity measures were selected from the biosecurity guidelines jointly published by MPI and PIANZ on the basis that they offer a good baseline for all producers. For this reason, it is more likely that broiler enterprises would have adopted the combination of biosecurity measures in the survey as the guidelines have been largely taken up by all the major poultry suppliers who govern over 90% of broiler producers (Muellner *et al.* 2016). Therefore, despite no legislation enforcing the adoption of biosecurity practices, the majority of broiler producers will be following the recommended guidelines under their company's policies on expected biosecurity practices in order to maintain a supply contract. In comparison to the broiler industry, the commercial layer industry consists of a much larger number of independent operators with only a few shared contractors. This means there is no common biosecurity policy setting, but instead, each individual producer is left to decide what measures they consider appropriate to mitigate on-farm risks, reducing the incentive to maintain a minimum set of biosecurity practices and resulting in a large amount of diversity between layer operations. Future surveys should expand on the range of biosecurity measures included

in the survey-questionnaire, or better still, provide a free-text response in which producers could add any biosecurity measures not found in the pre-selected list.

Our study found no significant relationship between the reported uptake of biosecurity measures and the potential level of transmission risk through contacts with a generally low biosecurity score across many farms. This is a concern given that highly connected farms can contribute disproportionately to disease transmission through contact networks especially if no control measures are in place to limit spread (Christley *et al.* 2005; Gates and Woolhouse, 2015). A similar lack in biosecurity has also been reported in a number of previous studies looking both at commercial poultry industries outside New Zealand (Dorea *et al.* 2010; Van Steenwinkel *et al.* 2011; Scott *et al.* 2018) and other livestock sectors (Gunn *et al.* 2008; Laanen *et al.* 2014) although it is not always clear why. In our study, free-text responses suggest that many producers have some level of concern or doubt over the effectiveness of biosecurity procedures, noting that they are often too costly or impractical to implement. However, it is also important to consider the high health status of New Zealand's national flock, which is free from major exotic avian diseases such as highly pathogenic avian influenza, Newcastle disease, and until recently infectious bursal disease (Davidson, 2002; Cobb and Smith, 2013), and where common pathogens like *campylobacter* have minimal economic impacts. This may be serving to reduce the producer's perception of risk both in terms of how vulnerable they feel towards a threat and how severe they think the potential consequences may be (Ferrer and Klein, 2015). Further research is needed to understand both the efficacy of different biosecurity measures in reducing disease spread and understanding how to motivate producers to increase biosecurity adoption.

There are several limitations in the study design that should be considered when interpreting the results. Given the low survey response rate, there is likely voluntary response bias particularly given that before completing the survey, all producers were

informed that the study objectives were to develop network simulation models in order to predict disease spread. Therefore, it is likely that those who chose to continue with the survey reflect producers that are more concerned about disease management and may practice different behaviours and practices to those producers who did not respond (Laanen *et al.* 2014). Reporting bias is also a further concern as results rely on the survey's ability to accurately reflect the reality on farm, however, erroneous reporting can happen for a number of reasons with many previous studies showing particularly poor correlation between survey responses and on-farm biosecurity practices (Sax *et al.* 2003; Bewsell, 2010; Racicot *et al.* 2012b). This reporting bias may be a result of pressure from the public, government, and other producers to maintain a high level of biosecurity; making it difficult for producers to speak openly about farm management practices in fear of potential repercussions and stigma. Further mismatches between the reported biosecurity and on-farm practices may also be because those completing the survey, often farm owners, may be unaware of employee practices or non-compliance overall, leading to an overestimation in the final biosecurity score.

In order to control some of these biases, future research focusing on capturing on-farm biosecurity and contact risk pathways may benefit from using mixed method approaches including qualitative methods, such as semi-structured interviews, to more accurately capture what is happening on farm. Farm visits may also increase the response rate of future surveys, particularly in the layer industry who have governing body to promote research participation. Survey fatigue was also quite clear with the majority of responders indicating that they would not be willing to participate in future surveys, however face-to-face farms visits allow the researcher to engage more with producers and discuss the importance of the research. Interviews also allow for a broader range of questions that could be useful to explore on-farm practices in more detail, especially given the current uncertainty around the different practices between different producers. In addition to qualitative approaches, the low response may be ameliorated in future studies by working

closer with industry who could not only help to endorse the survey but also have the power to make responses mandatory with PIANZ membership.

There were also limitations in the methods used to assess and quantify both the biosecurity and contact risk scores. Firstly, it was assumed that each biosecurity measure had equal importance. This assumption was made as there is very little research comparing the use of different biosecurity measures without reference to a specific disease, making it difficult to give weights or rank different measures. Nevertheless, it is important to consider that equal weighting may also be misleading and future studies may benefit by considering qualitative approaches, such as joint ranking, to avoid this assumption. Previous studies have also used expert opinion to give weights to biosecurity measures. For example, one study by Gelaude and colleagues (2014) was able to quantify the level of biosecurity on broiler farms by prioritizing and weighing various biosecurity measures taking into account the opinion of 16 different experts including epidemiologists, veterinary practitioners, and microbiologists (Gelaude *et al.* 2014). However, these weights depend on a number of different variables making it difficult to extrapolate a biosecurity scoring system across different studies (Sayer *et al.* 2014; Kuster *et al.* 2015).

The transmission risk score is similarly limited by the ability to accurately capture disease risk pathways using an indirect measure to assess the risk of disease transmission. In the current study, the contact risk pathways were selected from both a previous survey by Rawdon and colleagues (2007, 2008) and evidence-based in the literature, however, many of these pathways are complex and cannot be quantitatively assessed from a single observation (Rawdon *et al.* 2007; Rawdon *et al.* 2008). For example, the potential risk from contact with wild birds or waterfowl was measured by their presence or absence on-farm with no regard to variables such as the number of wild birds, their species and behaviour patterns such as migration and feeding traits, all of which are also thought to

play a role in the transmission dynamics between domestic and wild birds (Gilchrist, 2005; Spackman, 2009; Prosser et al., 2013).

Similarly to the biosecurity score, the relative importance of each pathway is also likely influenced by the characteristics of individual pathogens. Lastly, when comparing both the biosecurity and contact risk scores, it is important to consider the implications of using scoring systems based on a numerical range. For example, for the biosecurity score a value was given between zero and four for each biosecurity measure listed such that the difference between a farm with the minimum level of biosecurity and the optimum level could range from zero to 28. However, this score does not assume that the magnitude of change in biosecurity is directly proportional to the scale used in the scoring system (*i.e.*, a score of 28 does not imply that the biosecurity is double that for a score of 14). The small range between the minimum and optimum biosecurity score also results in many of the study farms being grouped together, implying that the level of biosecurity between the farms is very similar. This emphasises need to develop reliable and accessible tools that can be used to assess both on-farm risks and biosecurity practices as well as providing a benchmark to help individuals maintain a high level of biosecurity. However, despite these limitations in capturing and quantifying both on-farm biosecurity and contact risk pathways, it is still clear that there is a large variation between commercial poultry farms in New Zealand.

3.6. Conclusion

The study findings highlight the diversity in contact risk pathways and biosecurity practices across a subset of farms in New Zealand the commercial poultry industry. From a disease control perspective, it is concerning that farms with the highest potential level of risk for acquiring and/or spreading disease through the contact network were no more or less likely to adopt biosecurity measures to prevent disease transmission than farms with relatively few contacts. This may be related to the low prevalence or absence of many

important infectious poultry diseases in New Zealand, leading farmers to believe there is limited need to maintain good biosecurity as well as farmer uncertainty around the efficacy of different biosecurity measures. Further research is needed to understand (i) how producers are identifying and assessing disease risks and (ii) which factors are most important in motivating long-term changes in risk management behaviours.

3.7. Acknowledgements

The authors are grateful to the Poultry Industry Association of New Zealand and the Egg Producers' Federation of New Zealand for their support in administering the survey to poultry producers and further thanks to both Ahmed Fayaz and Shekhar Pokhrel for their assistance in managing the database and data entry. Funding for this research was provided by the Ministry for Primary Industries, the School of Veterinary Sciences (Massey University), and New Zealand's Biological Heritage: National Science Challenges.

3.8. References

- Akintunde OK, Adeoti AI. Assessment of factors affecting the level of poultry disease management in southwest, Nigeria. *Trends Agri. Econ.* 7, 41–56, 2014
- Anonymous. *Primary production – poultry: June 2018 quarter and year – Infoshare tables.* Statistics New Zealand. Available at: <https://www.stats.govt.nz/information-releases/primary-production-poultry-june-2018-quarter-and-year-infoshare-tables>. Accessed 26 August 2019. Stats NZ, New Zealand, 2018
- Anonymous. *PIANZ guidelines for the use of antibiotics in poultry.* Poultry Industry Association of New Zealand (PIANZ). Available at: <https://www.pianz.org.nz/news/pianz-guidelines-for-the-use-of-antibiotics-in-poultry/>. Accessed 01 November 2019. PIANZ, Auckland, New Zealand, 2017
- Astill J, Fraser E, Dara R, Sharif S. Detecting and predicting emerging disease in poultry with the implementation of new technologies and big data: a focus on avian influenza virus. *Front. Vet. Sci.* 5, 263–275, 2018
- Bewsell D. *Determining individuals' response to New Zealand biosecurity.* Degree of Masters of Commerce and Management, Available at: <https://pdfs.semanticscholar.org/1e8e/6160eb8253e90bd98921d2585db5336766f3.pdf> Lincoln University, Lincoln, New Zealand, 2010
- Bridges VE, Akkina J, Grannis J, Johnson C, Johnson R, Tuszynski C. A qualitative assessment tool for the potential of infectious disease emergence and spread. *Prev. Vet. Med.* 81, 80–91, 2007
- Christley RM, Pinchbeck GL, Bowers RG, Clancy D, French NP, Bennett R, Turner J. Infection in social networks: using network analysis to identify high-risk individuals. *Am. J. Epidemiol.* 162, 1024–1031, 2005
- Cobb S, Smith H. *Import risk analysis: chicken and duck meat for human consumption.* Ministry for Primary Industries, Wellington, New Zealand. Available at: <https://www.mpi.govt.nz/dmsdocument/2815>. Accessed on 26 October 2019. MPI, 2013
- Davidson RM. Control and eradication of animal diseases in New Zealand. *N. Z. Vet. J.* 50, 6–12, 2002
- Dorea FC, Berghaus R, Hofacre C, Cole DJ. Survey of biosecurity protocols and practices adopted by growers on commercial poultry farms in Georgia, U. S. A. *Avian Dis.* 54, 1007–1015, 2010

- Dunn OJ.** Multiple comparison using RANK sums. *Technometrics* 6, 241–252, 1964
- East I, Kite V, Daniels P, Garner G.** A cross-sectional survey of Australian chicken farms to identify risk factors associated with seropositivity to Newcastle-disease virus. *Prev. Vet. Med.* 77, 199–214, 2006
- Ferrer RA, Klein WM.** Risk perceptions and health behaviour. *Curr. Opin. Psychol.* 5, 85–89, 2015
- Fraser RW, Williams NT, Powell LF, Cook AJC.** Reducing *Campylobacter* and *Salmonella* infection: two studies of the economic cost and attitude to adoption of on-farm biosecurity measures. *Zoonoses Public Health* 57, 109–115, 2010
- Fruchterman TMJ, Reingold EM.** Graph drawing by force-directed placement. *Software Pract. Exp.* 21, 1129–1164, 1991
- Funk S, Salathé M, Jansen VAA.** Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J. R. Soc. Interface* 7, 1247–1256, 2010
- Garforth CJ, Bailey AP, Tranter RB.** Farmers’ attitudes to disease risk management in England: a comparative analysis of sheep and pig farmers. *Prev. Vet. Med.* 110, 456–466, 2013
- Gates MC, Woolhouse ME.** Controlling infectious disease through the targeted manipulation of contact network structure. *Epidemics* 12, 11–19, 2015
- Gelaude P, Schlepers M, Verlinden M, Laanen M, Dewulf J.** BiocheckUGent: a quantitative tool to measure biosecurity at broiler farms and the relationship with technical performances and antimicrobial use. *Poultry Sci.* 93, 2740–2751, 2014
- Gilchrist P.** Involvement of free-flying wild birds in the spread of the viruses of avian influenza, Newcastle disease and infectious bursal disease from poultry products to commercial poultry. *World Poultry Sci. J.* 61, 198–214, 2005
- Gunn GJ, Heffernan C, Hall M, McLeod A, Hovi M.** Measuring and comparing constraints to improved biosecurity amongst GB farmers, veterinarians and the auxiliary industries. *Prev. Vet. Med.* 84, 310–323, 2008
- Hidano A, Enticott G, Christley RM, Gates MC.** Modeling dynamic human behavioural changes in animal disease models: challenges and opportunities for addressing bias. *Front. Vet. Sci.* 5, 137–151, 2018

- Itagaki K.** *Evaluating the role of public opinion in the New Zealand Biosecurity Act 1993.* Degree of Master of Resource Studies. Available at: <https://researcharchive.lincoln.ac.nz/handle/10182/5703>. Lincoln University, Lincoln, New Zealand, 2013
- Kruger H, Thompson L, Clarke R, Stenekes N, Carr A.** *Engaging in biosecurity: gap analysis.* Australian Government Bureau of Rural Sciences. Available at: <http://www.agriculture.gov.au/SiteCollectionDocuments/abares/publications/biosecuritygapsanalysis.pdf>. Accessed on 26 October 2019. BSR, Canberra, Australia, 2009
- Kuster K, Cousin M-E, Jemmi T, Schüpbach-Regula G, Magouras I.** Expert opinion on the perceived effectiveness and importance of on-farm biosecurity measures for cattle and swine farms in Switzerland. *PLoS One* 10, e0144533, 2015
- Laanen M, Maes D, Hendriksen C, Gelaude P, De Vlieghe S, Rosseel Y, Dewulf J.** Pig, cattle and poultry farmers with a known interest in research have comparable perspectives on disease prevention and on-farm biosecurity. *Prev. Vet. Med.* 115, 1–9, 2014
- Lestari VS, Natsir A, Sirajuddin SN, Kasim K, Ali HM, Saadah, Mawardi.** Factors influencing biosecurity adoption on laying hen farmers. *J. Indonesian Trop. Anim. Agric.* 37, 302–307, 2012
- Lewerin SS, Österberg J, Alenius S, Elvander M, Fellström C, Tråvén M, Wallgren P, Waller KP, Jacobson M.** Risk assessment as a tool for improving external biosecurity at farm level. *BMC Vet. Res.* 11, 171–183, 2015
- Lockhart CY, Stevenson MA, Rawdon TG, Gerber N, French NP.** Patterns of contact within the New Zealand poultry industry. *Prev. Vet. Med.* 95, 258–266, 2010
- Mankad A.** Psychological influences on biosecurity control and farmer decision-making. A review. *Agron. Sustain. Dev.* 36, 40, 2016
- Martínez-López B, Perez AM, Sánchez-Vizcaíno JM.** Combined application of social network and cluster detection analyses for temporal-spatial characterization of animal movements in Salamanca, Spain. *Prev. Vet. Med.* 91, 29–38, 2009
- Moore DA, Merryman ML, Hartman ML, Klingborg DJ.** Comparison of published recommendations regarding biosecurity practices for various production animal species and classes. *J. Am. Vet. Med. Assoc.* 233, 249–256, 2008

- Muellner P, Kells N, Campbell D. *Risk profile: the emergence of Campylobacter jejuni ST 6964 in poultry in New Zealand and its associated antimicrobial resistance*. Available at: <https://www.epi-interactive.com/sites/default/files/documents/2016-16-Risk-Profile-Campylobacter-ST6964-with-SIS.PDF>. Accessed 24 September 2018. Ministry for Primary Industries Technical Paper No: 2016/16, Wellington, New Zealand, 2016
- Müllner P, Collins-Emerson JM, Midwinter AC, Carter P, Spencer SE, van der Logt P, Hathaway S, French NP. Molecular epidemiology of *Campylobacter jejuni* in a geographically isolated country with a uniquely structured poultry industry. *Appl. Environ. Microbiol.* 76, 2145-2154, 2010
- Nespeca R, Vaillancourt J-P, Morgan Morrow WE. Validation of a poultry biosecurity survey. *Prev. Vet. Med.* 31, 73–86, 1997
- Niemi JK, Lyytikäinen T, Sahlstrom L, Virtanen T, Lehtonen H. *Risk classification in animal disease prevention: who benefits from differentiated policy?* Proceedings for the 2009 Annual Meeting of Agricultural and Applied Economics Association. Available at: <https://ideas.repec.org/p/ags/aaea09/49307.html>. Agricultural and Applied Economics Association, Milwaukee, Wisconsin, 2009
- Olson CK, Ethelberg S, Pelt W van, Tauxe RV. Epidemiology of *Campylobacter jejuni* infections in industrialized nations. In: *Campylobacter*. Pg. 163–189. ASM Press, Washington, DC, 2008
- Prosser DJ, Hungerford LL, Erwin RM, Ottinger MA, Takekawa JY, Ellis EC. Mapping avian influenza transmission risk at the interface of domestic poultry and wild birds. *Front. Public Health* 1, 28-39, 2013
- Poultry Industry Association of New Zealand (PIANZ). *Biosecurity manual for New Zealand meat chicken growers*. Available at: <https://pianz.org.nz/wp-content/uploads/2015/11/biosecurity-manual-september-2015.pdf>. Accessed 24 September 2018. PIANZ, Auckland, New Zealand, 2015
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.R-project.org/>. 2010
- Racicot M, Venne D, Durivage A, Vaillancourt J-P. Evaluation of the relationship between personality traits, experience, education and biosecurity compliance on poultry farms in Québec, Canada. *Prev. Vet. Med.* 103, 201–207, 2012a

- Racicot M, Venne D, Durivage A, Vaillancourt J-P.** Evaluation of strategies to enhance biosecurity compliance on poultry farms in Québec: Effect of audits and cameras. *Prev. Vet. Med.* 103, 208–218, 2012b
- Rautureau S, Dufour B, Durand B.** Structural vulnerability of the French swine industry trade network to the spread of infectious diseases. *Animal* 6, 1152–1162, 2012
- Rawdon T, Tana T, Frazer J, Thornton R, Chrystal N.** Biosecurity risk pathways in the commercial poultry industry: free-range layers, pullet-rearers and turkey broilers. *Surveillance* 35, 4–9, 2008
- Rawdon T, Thornton R, McKenzie J, Gerber N.** Biosecurity risk pathways in New Zealand's commercial poultry industry. *Surveillance* 34, 4–9, 2007
- Sánchez-Matamoros A, Martínez-López B, Sánchez-Vizcaíno F, Sánchez-Vizcaíno JM.** Social network analysis of Equidae movements and its application to risk-based surveillance and to control of spread of potential Equidae diseases. *Transbound. Emerg. Dis.* 60, 448–459, 2013
- Sax LJ, Gilmartin SK, Bryant AN.** Assessing response rates and nonresponse bias in web and paper surveys. *Res. High. Educ.* 44, 409–432, 2003
- Sayers RG, Good M, Sayers GP.** A survey of biosecurity-related practices, opinions and communications across dairy farm veterinarians and advisors. *Vet. J.* 200, 261–269, 2014
- Scott AB, Singh M, Groves P, Hernandez-Jover M, Barnes B, Glass K, Moloney B, Black A, Toribio J-A.** Biosecurity practices on Australian commercial layer and meat chicken farms: Performance and perceptions of farmers. *PLoS One* 13, e0195582, 2018
- Spackman E.** The ecology of avian influenza virus in wild birds: What does this mean for poultry? *Poultry Sci.* 88, 847–850, 2009
- Susilowati SH, Patrick I, Iqbal M, Jubb T.** The characteristics of the farm and the farmer that affect the adoption of biosecurity on smallholder poultry farms in Indonesia. *Livestock Res. Rural Dev.* 25, 2013
- Taveirne ME, Dunham DT, Perault A, Beauchamp JM, Huynh S, Parker CT, DiRita VJ.** Complete annotated genome sequences of three *Campylobacter jejuni* strains isolated from naturally colonized farm-raised chickens. *Genome Announc.* 5, 2017

van Steenwinkel S, Ribbens S, Ducheyne E, Goossens E, Dewulf J. Assessing biosecurity practices, movements and densities of poultry sites across Belgium, resulting in different farm risk-groups for infectious disease introduction and spread. *Prev. Vet. Med.* 98, 259–270, 2011

Zhang M, He L, Li Q, Sun H, Gu Y, You Y, Meng F, Zhang J. Genomic characterization of the Guillain-Barre syndrome-associated *Campylobacter jejuni* ICDCCJ07001 Isolate. *PLoS One* 5, e15060, 2010

**Characterising the disease risk from backyard
poultry trade networks and migratory birds to
New Zealand's commercial poultry industry**

Sabrina S. Greening, Kerry Mulqueen, Thomas G. Rawdon, Nigel P. French, and
M. Carolyn Gates

4.1. Abstract

AIMS: Commercial poultry farms are at risk of disease introductions through potential contacts with both backyard poultry producers and wild bird populations. However, it has historically been difficult to estimate this risk due to the lack of accurate data on the numbers, locations, and movement patterns of these three populations. In this study, we explore the use of different data sources to determine the disease transmission risk within the New Zealand commercial poultry industry from (i) the sale of backyard poultry through an online auction website and (ii) the movement of wild birds both within New Zealand and from overseas.

METHODS: Results from a cross-sectional survey conducted in 2016 of all registered New Zealand commercial poultry operations were used to characterise contact patterns within the industry and investigate potential contacts with both backyard poultry producers and wild bird populations. For backyard poultry, a data extract containing all recorded poultry transactions made through the online auction website TradeMe® from 2012 to 2018 were used to construct two poultry trade networks characterising the connectivity of individual traders and spatial locations. The spatial network was then used to create a Susceptible-Infectious network simulation model to help characterise the potential for an epidemic disease to spread across New Zealand through the movement of backyard poultry. To determine the disease transmission risk from migrating wild birds, both within New Zealand and from overseas, a data extract from the online citizen science bird monitoring project eBird was analysed. Bird observations made from 2012 to 2018 were used to characterise the spatial distribution of wild birds. The overlap between commercial poultry premises, backyard poultry trade, migratory birds and water birds belonging to the Anatidae family was explored using bivariate choropleth maps.

RESULTS: Analysis of the online auction data revealed highly active backyard poultry trade networks with a total of 137,270 recorded trades between 59,225 traders during the

study time period. There was a high level of connectivity and strong spatial linkages between the major urban centres that followed a similar pattern to that seen in the commercial poultry network suggesting significant overlap between backyard poultry trade and commercial poultry. Results from the simulation model shows the potential for disease to spread through these trades with all the 134 suburb nodes becoming infected in 96.4% (9,642/10,000) of the simulations. Analysis of the eBird data included 73,990 reports sighting 80 bird species known to migrate to, from or within New Zealand. The majority of these migratory bird species belonged to the order Charadriiformes (coastal seabirds and wading birds) and posed little risk to commercial poultry with high numbers reported in isolated coastal areas. Resident birds that migrate within New Zealand and water birds were of greatest concern with many reported in the same habitat as both exotic migratory birds and inland in regions with a high density of commercial and backyard poultry.

CONCLUSIONS: Overall, our study findings highlight how the spatial patterns of online poultry trade and migratory birds can influence the risk landscape within the commercial poultry industry. In particular, the high volume of animals traded through online auction websites over increasingly long distances and shorter timespans will have important implications for disease transmission dynamics. The significant overlap between wild bird populations and backyard poultry also increases the risk of disease introduction and spread, particularly given the generally low standards of biosecurity for backyard poultry. To reduce this risk, it is essential that future disease prevention and control strategies consider increasing biosecurity education and regulations within the backyard poultry sector.

KEY WORDS: Backyard poultry, Migratory wild birds, Contact networks, Disease transmission risk

4.2. Introduction

Controlling the spread of infectious disease through poultry and livestock production systems requires detailed knowledge about the location of farms and the patterns of contact between them (Garner *et al.*2007). As such, most developed countries have established national databases that require farmers to provide up-to-date information on the numbers of animals in their care as well as the movements of animals between registered locations for trade and other purposes such as grazing, breeding, veterinary care, or exhibition at shows (Saatkamp *et al.*1995). This had led to a significant improvement in the quality of data in both commercial poultry and livestock operations. However, it is widely recognized that these systems often fail to capture information on non-commercial or “backyard” producers, defined as those individuals who keep a small number of animals for personal consumption or as a hobby (Johnson *et al.*2004). This knowledge gap presents a major concern for commercial operations and health authorities as there is a potential for backyard producers to act as major disease reservoirs, increasing the risk of disease transmission to not only commercial operations but also the general public if the disease is zoonotic (Behravesch *et al.*2014; Pohjola *et al.*2016).

In an attempt to mitigate this risk, there is a limited number of studies that have tried to characterise the risk of disease from backyard poultry to commercial poultry producers with a focus on the local spread of endemic diseases where backyard poultry are in close contact with both commercial poultry and wild birds or waterfowl (Johnson *et al.*2004; Derksen *et al.*2008; Fiebig *et al.*2009). Further studies have also investigated the general structure of the backyard poultry sector in terms of its size and composition, the geographical distribution of households keeping poultry, and their knowledge and management practices of disease control in order to help assess the risk backyard poultry pose to the commercial poultry industry (Jutzi, 2005).

More recently, with improvements in our ability to track wild populations, an increasing number of studies have also investigated the risk of exotic disease being introduced into commercial poultry via direct contact with migratory bird populations but also via contact with backyard poultry that have also been exposed to migratory birds (Hamilton-West *et al.* 2012; Wang *et al.* 2013). For disease introduction, the latter pathway is of growing concern with a rapid expansion in the number of households keeping poultry (USDA, 2013) in conjunction with the lack of regulations aimed at the backyard poultry sector and the varying levels of knowledge between backyard producers that often result in less than optimal levels of biosecurity (Smith and Dunipace, 2011; Schembri *et al.* 2015). However, in order to fully characterise these risks, studies must be able to accurately capture information on both backyard poultry and wild birds as well as being able to determine if any contact pathways exist between these populations and commercial poultry enterprises.

In New Zealand, a number of studies have used a range of data sources to characterise the risk from backyard poultry to the commercial poultry industry. For example, Zheng and colleagues (2010) used a traditional cross-sectional survey data and diagnostic samples from 54 non-commercial poultry owners to identify several possible transmission pathways that posed a risk of spreading avian influenza (AI) to commercial operations with survey results highlighting the potential for backyard poultry to be exposed to AI through direct contact with wild birds (Zheng *et al.* 2010). This is in keeping with a further study by Lockhart and colleagues (2010) that also used a cross-sectional survey to record details from a subset of backyard poultry producers in urban and peri-urban settings within New Zealand, with results also highlighting the importance of strict biosecurity measures on commercial farms due to the close proximity of backyard poultry with a low level of biosecurity (Lockhart *et al.* 2010a).

However, in both of these previous studies, it was recognised that the data most likely failed to capture a large proportion of contacts involving backyard producers due to the limitations in both the geographic region and/or time scale over which the data was collected. Similar limitations have been seen in studies starting to utilise less traditional data sources in an attempt to infer information on the demographic and contact patterns of backyard producers such as the registration data at poultry shows (Hernández-Jover *et al.* 2013; Hernández-Jover *et al.* 2015). These alternative data sources offer opportunities to re-evaluate the risks posed by backyard producers. For example, in New Zealand, a popular method to buy and sell both poultry and livestock is through the online trading website TradeMe® (www.trademe.co.nz) and although it is not possible to determine what fraction of all backyard poultry trades occur through this site, it is anecdotally believed to be the most popular marketing channel, with over 3 million registered users out of a total country population of 4.2 million, representing a significant number of backyard poultry movements that have not been accounted for.

Data sources regarding the spatial distribution and movement patterns of wild birds is far more limited and although data and information on migratory and resident birds is frequently collected for research, management and conservation, it is often restricted to only a small number of species within a single geographical area (Isaac *et al.* 2014). This is the case with several surveys in New Zealand that have looked at wild bird populations many of which have been a part of the national surveillance program for AI, and therefore have been focused on those species considered high risk (Tana *et al.* 2007; Frazer *et al.* 2008-2010; Langstaff *et al.* 2009). More recently, studies have turned to citizen science projects as alternative data sources that can be used to help determine the distribution of different species. A citizen scientist is any member of the general public that collects and analyses data typically as part of a collaborative project with a scientific team. Despite the generated data often being less structured with a high amount of variability and bias, it is high in quantity and has allowed the scientific community to address many questions on

broad temporal and spatial scales that would otherwise be logistically or financially unfeasible (Dickinson *et al.*2010; Bird *et al.*2014). One such project looking at the abundance and distribution of wild birds is eBird (<https://ebird.org>). Launched in 2002 by the Cornell Lab of Ornithology at Cornell University and the National Audubon Society, eBird is the world's largest biodiversity-related citizen science project and although originally recordings were restricted to the Western Hemisphere, the project was expanded in 2008 to include New Zealand (<https://ebird.org/newzealand>) and then again in 2010 for worldwide coverage. Since its launch over 420,000 people have participated in the project with more than 590 million bird observations in the database contributing to over 220 peer-reviewed scientific publications (eBird, 2018).

The main objectives of this study were to explore the use of different data sources to (i) characterise the trading network of backyard poultry sold through the online auction website TradeMe®, and (ii) determine the spatial pattern of wild birds that are known to migrate to, from and within New Zealand using eBird observations, (iii) investigate the risk of diseases spreading between non-commercial poultry, backyard poultry and wild birds populations and, (iv) evaluate the use of these datasets in supporting disease preparedness and response efforts.

4.3. Materials and Methods

4.3.1. Commercial poultry network

Contact details for commercial poultry enterprises in New Zealand were obtained from the Poultry Industry Association of New Zealand (PIANZ) and the Egg Producers' Federation (EPF). The database was accessed in June 2016 and listed 426 enterprises including hatcheries, breeding and rearing units, layer farms and broiler production units. The addresses provided in the database were used to retrieve longitude and latitude co-ordinates using the R package *ggmap* (Kahle and Wickham, 2013). The co-ordinates were then plotted to visualise the spatial distribution of poultry enterprises across New

Zealand. Additional choropleth maps were constructed using the R package *ggplot2* (Wickham, 2016) to identify the territorial authorities (TAs) with the highest density of commercial premises for comparison with a map showing the spatial distribution of migratory birds described below.

In addition, a contact network was constructed to show all on- and off-farm movements relating to feed, waste, litter, live birds, hatching eggs, table eggs and additional poultry products *i.e.*, offal and feathers. The reported movements were obtained from the results of an industry survey administered to all active poultry producers in New Zealand. The survey was based on a previous questionnaire conducted by Lockhart and colleagues in 2006 (Lockhart *et al.* 2010b) and modified in collaboration with PIANZ, EPF and the Ministry for Primary Industries (MPI) with an aim of collecting information on the farm demographics, contact patterns, and biosecurity practices of New Zealand commercial poultry operations. The study was judged to be low risk thorough peer evaluation and consequently was not formally reviewed by any of the University's Human Ethics Committees. Full details on the survey design and implementation have been described in Chapter 3 and a copy of the complete survey questionnaire is available in Appendix A. In the network, nodes represented commercial poultry premises that responded to the industry survey with undirected edges linking nodes who reported the use of a common company delivery goods or services on- and off-farm (Figure 4.1). A network graph was constructed using a force-based algorithm proposed by Fruchterman and Reingold (1991) in the R package *igraph* (Csardi and Nepusz, 2006). Basic network statistics such as measures of centrality and cohesion have been described in Chapter 3, while for this study the network graph was plotted onto a geographical map so that nodes were positioned on their corresponding farm coordinates. The plot was created using the R package *ggplot2* (Wickham, 2016) for comparison with the spatial networks constructed from the sale of backyard poultry described below.

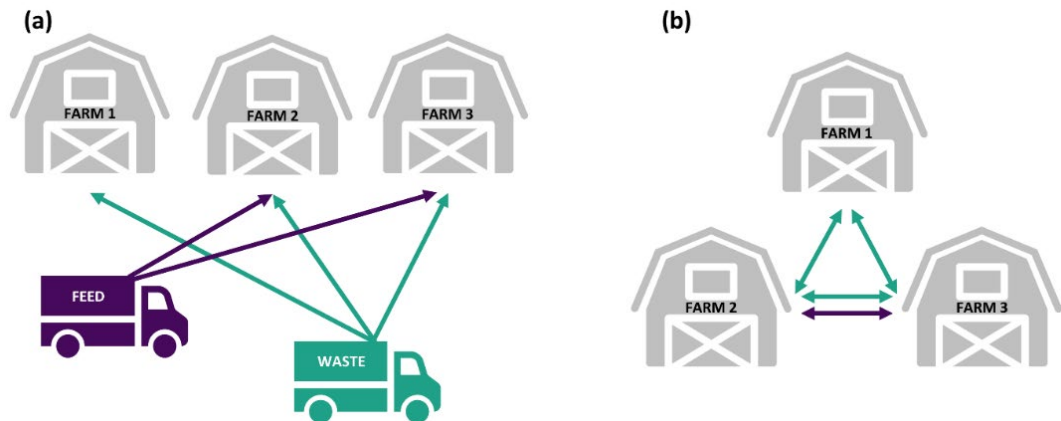


Figure 4.1. Schematic diagram showing the construction of the commercial poultry network with (a) showing three farms connected to transporting vehicles belonging to different companies: one providing feed, the other removing waste. All three farms use the waste company (blue lines) whilst only Farms 2 and 3 use the feed company (orange lines) with those farms linked to the same company being directly linked in (b) such that farms 1 is connected to farms 2 and 3 via the movement of waste, and farms 2 and 3 are connected via the both the movement of feed and waste.

4.3.2. Backyard poultry network

4.3.2.1. TradeMe® transactions

A data extract containing records of all online poultry auction sales occurring through the TradeMe® website over a seven-year period from 01st January 2012 to 31st December 2018 was provided. This included details on the (i) trade date, (ii) anonymized seller and buyer identification numbers, (iii) geographic location of the seller and buyer at the region and suburb levels, (iv) type of poultry traded (bantams, chickens, and ducks), (v) number of animals traded and, (vi) free-text descriptions provided by the seller on the poultry for sale. It should be noted that the number of animals traded represents the lower bounds since there were many trades that had “1” recorded under number of animals traded, but where the free-text descriptions indicated that multiple animals were being sold. Hereafter, we collectively referred to sellers and buyers registered on the TradeMe® website as “traders”. Trades to Australia and the Chatham Islands were excluded from the analysis as they were considered to be a low disease risk. Descriptive statistics were

calculated summarising the number and frequency of transactions and the number of active traders each year during the study time period, stratified by poultry type. All analyses were performed using the R statistical software (R Core Team, 2018).

4.3.2.2. Backyard poultry networks

For each year, two separate networks were constructed: (i) with nodes representing the traders and (ii) with nodes representing suburbs. The trader network was used to characterise patterns in individual connectivity whilst the suburb network was used to characterise patterns in spatial connectivity. Data from the 2018 calendar year (01st January 2018 to 31st December 2018) have been used as an exemplar year. Network graphs were constructed using the R package *igraph* (Csardi and Nepusz, 2006) with links between nodes representing a poultry transaction for all poultry types (chickens, bantams and ducks) in both the trader networks and spatial networks. The network statistics, described in Table 4.1, were calculated for each network including node in- and out-degree, betweenness, density, diameter, average path length, network clustering coefficient, number of strongly connected and weakly connected components (SCC and WCC respectively), the size of the giant strongly connected component (GSCC) and the giant weakly connected component (GWCC), network reciprocity, and network fragmentation. Degree distributions for the number of inward contacts, number of outward contacts, and the total number of contacts, were also calculated for the trader network to determine if the scale-free property in which a small number of individuals contribute a disproportionately large number of links in the network (Barabási, 2009) existed as is common in other commercial livestock networks (Hardstaff *et al.* 2015).

To visualize the spatial pattern of poultry sales and identify regions with a large amount of trade, the spatial networks were plotted onto geographical maps using the R package *ggplot2* (Wickham, 2016) in which nodes were positioned on the regional centroids corresponding to the geographic location of the trader as identified in the database. Three

additional network graphs were constructed for each year to compare and contrast the networks constructed from the transaction of the different poultry types. In each graph the nodes represented the traders whilst the links between nodes represented the transaction of either chickens, bantams, or ducks.

Table 4.1. Network analysis glossary of terms used to describe the trader networks and spatial networks representing poultry auction sales occurring through the TradeMe^{*} website in New Zealand.

Network statistic	Definition
<i>Density</i>	The proportion of all possible links between nodes in the network that are present
<i>Diameter</i>	The longest path between any two pair of nodes in the network
<i>In-degree</i>	Number of individual trades received by a node in the network
<i>Out-degree</i>	Number of individual trades dispatched by a node in the network
<i>Total-Degree</i>	The sum of the in- and out-degree
<i>Betweenness</i>	The frequency a node is found on the shortest path between any other two pair of nodes in the network
<i>Average path length</i>	The average shortest path between any pair of nodes in the network averaged over all pairs of nodes
<i>Clustering coefficient</i>	For any node in the network the clustering coefficient is the proportion of neighbouring nodes in direct contact with the node that are also connected to each other.
CC	CC: Connected Components. A set of nodes within the network in which all nodes are mutually accessible. In a strongly connected component (SCC) each pair of nodes in the connected component is reachable by following the direction of the link whereas in a weakly connect component (WCC) the direction of the link is ignored.
GSCC	GSCC: Giant Strongly Connected Component. The largest strongly connected component (SCC) in a directed network in which all nodes are mutually accessible by following the direction of the link.
GWCC	GWCC: Giant Weakly Connected Component. The largest weakly connected component (WCC) in the network that is an undirected network in which all nodes are linked.
<i>Reciprocity</i>	The likelihood of nodes to be mutually linked <i>i.e.</i> , the likelihood that any pair of nodes in the network both receive and dispatch trades between each other.
<i>Fragmentation</i>	The proportion of node pairs for which a path does not exist between them.

4.3.2.3. Backyard poultry disease simulation model

To explore the potential for an epidemic disease, such as avian influenza, to spread across New Zealand through the movement of backyard poultry, a Susceptible-Infectious (SI) network simulation model was created modelling poultry trades between the 134 suburb nodes found in the largest of the spatial networks. Given the limited data to inform within-suburb transmission dynamics, the objective of this model was to reproduce the “worst case” scenario assuming maximum transmission between suburbs; such that, if a movement occurred between an infected suburb to a susceptible suburb the probability of transmission was fixed at one with no delay between the time in which a transmission event occurred and the time it takes for the entire suburb to be infected. Using this approach network structures that support worst-case scenarios can be identified, an exercise that has been applied to many diseases, such as smallpox (Kaplan *et al.* 2002), measles, norovirus (Daughton *et al.* 2017) and influenza (Keeling and Danon, 2009), for which the available data precludes accurate prediction of disease spread but models have still been important decision support tools aiding public health planning.

For each model simulation, a single suburb was selected as the primary infectious case on any random day between 01st January 2012 and 31st December 2018. The simulation was then updated on a daily basis with any trades from an infectious source suburb causing the destination suburb to become immediately infectious. The explicit patterns of daily transactions recorded in the data set were used to model the connections between suburb nodes. The simulation was stopped either when all the suburb nodes were infected or when a maximum of seven years had elapsed. If the simulation reached the calendar date of 31st December 2018 before either of those conditions were met, we recycled the data starting from 01st January 2012 with a total of 10,000 simulations performed. For each simulation, the following information was recorded: (i) the date of seeding, (ii) the suburb in which the disease was seeded, (iii) whether or not an individual suburb node became infected during the run, and (iv) the number of days from the start of the outbreak until

each individual suburb node became infected. For each individual suburb node, descriptive statistics were generated on the mean and median number of days it took for the suburb node to become infected across the 10,000 simulation runs.

4.3.3. Migrating wilds birds and waterfowl

Bird observation data was extracted from the citizen science project, eBird (Sullivan *et al.* 2009); an online bird monitoring project in which participants report the time, location, search effort and count of birds in a standardized manner. New Zealand presence-absence checklist data from 01st January 2012 to 31st December 2018 was requested for this analysis, with access granted on the 04th April 2019. Over 1 million observations had been recorded, however for the purpose of this analysis only records that could identify the bird species was used. To begin with, a choropleth map showing the total number of bird species reported within each TA in each year was created as a proxy measure for bird population and diversity. Those bird species not classified as migratory birds were then excluded in further analyses. A complete list of bird species known to migrate to, from or within New Zealand was obtained from the online encyclopaedia of New Zealand birds (<http://nzbirdsonline.org.nz>). For this analysis, it was assumed that seabirds that are known to spend large periods of time flying at sea, returning to land only for breeding, would be a low disease risk and therefore were excluded including all albatrosses (Diomedidae family), petrels (Oceanitidae family), skuas (Stercorariidae family) and gannets (Sulidae family). The remaining migratory bird species were cross referenced with the eBird data extract using both their common and scientific names to exclude non-migratory birds from the dataset. The remaining records were then categorised into resident bird species that only migrate within New Zealand and migratory bird species that migrate to and from New Zealand. These categories were used to distinguish between those bird species that present more of a risk for spreading endemic diseases within New Zealand versus bird species that present a risk of introducing an exotic disease from overseas respectively. Records observed on the islands

surrounding New Zealand coastline including the Auckland Islands, Snares Islands, Chatham Island, Pitt Island, Campbell Island and Antipodes Island were further excluded as these observations were also assumed to have little impact on disease risk.

In addition to the migratory bird species described above, all birds belonging to the family Anatidae, a biological family of water birds that includes ducks, geese, and swans, that do not migrate but are known to be found in New Zealand were also identified in the online encyclopaedia of New Zealand birds (<http://nzbirdsonline.org.nz>) and cross-referenced with the eBird data extract. This family of birds was included in the analysis as they are recognised as one of the largest natural host reservoirs for low pathogenic AI (Webster *et al.* 1992) that can not only result in severe disease and fatality in chickens (Short *et al.* 2015) but also presents a major risk for the introduction of highly pathogenic avian influenza virus H5N1 (HPAI) into New Zealand. Currently, the H5N1 virus has been detected in many parts of Asia, Europe and Africa with over 800 cases and 400 deaths in humans reported since 2003 and millions of deaths in domestic poultry and wild birds as a result from both the effects of the disease and culling efforts (Alexander, 2007; Whitworth *et al.* 2007; WHO, 2019). However, the H5N1 virus has yet to be detected in New Zealand although a number of species within the Anatidae family have been previously found positive for low pathogenic avian influenza, including those of the H5 or H7 subtype (Rawdon *et al.* 2007) that have the potential to evolve into HPAI viruses (Moone *et al.* 2014).

To visualize the spatial distribution of the migratory birds, both resident and exotic, and the water birds in the eBird data extracts, longitude and latitude coordinates were used to plot a map showing the location at which every observation was recorded in each given year over the seven-year study period. Choropleth maps were constructed using the R package *ggplot2* (Wickham, 2016) to identify regions with a high number of observations. To account for higher detection rates in urban centres with larger populations, maps were

adjusted by the population density within each TA using population count data from the 2013 census accessed via StatsNZ (<https://www.stats.govt.nz>).

4.3.4. Disease transmission risk between commercial and non-commercial poultry

To identify regions with a high chance of contact between commercial poultry and either backyard poultry or wild birds, a range of bivariate choropleth maps were constructed highlighting those TAs with both a high number of commercial poultry enterprises and (i) backyard poultry trades, (ii) bird species migrating to and from New Zealand, (iii) bird species migrating only within New Zealand, and (iv) bird species belonging to the Anatidae family. Additional bivariate choropleth maps were also constructed to identify regions with a high number of backyard poultry trades and (i) bird species migrating to and from New Zealand, (ii) bird species migrating only within New Zealand, and (iii) bird species belonging to the Anatidae family.

Within each TA the number of commercial poultry premises and bird observations were calculated using the point data described above including all the poultry producers registered in the PIANZ-EPF database (*i.e.*, both survey responders and non-responders). For backyard poultry trades, the spatial network was used to calculate an annual degree measure for each TA; that is the total number of trades made to, from or within the region over a one-year period. For each bivariate map a 9-class sequential colour scheme with each TA being categorized in to three classes: low, medium and high such that when variables were combined (*i.e.*, the number of commercial poultry enterprises, the annual trading degree or the number of wild bird observations) all combination resulted in a 3x3 grid that could be reflected using the 9-class sequential colour scheme. The cut-off points within each variable were calculated by dividing the range into terciles, creating the three classes. TAs could then be grouped by tercile with those in the lowest tercile considered to have low values relative to the other observations, those in the middle tercile having medium values and those in the highest tercile considered to have high values.

4.4. Results

4.4.1. Commercial poultry network

Overall, 414 active commercial poultry premises were identified in the PIANZ-EPF database with a high density of productions premises located within Auckland, New Plymouth, Selwyn, Waikato, Matamata-Piako and Waimakariri (Figure 4.2). As of October 2016, 29.0% (120/414) of the poultry producers had returned a completed postal survey with responses varying by production type such that broiler enterprises constituted 47.5% (57/120) of the survey responses whilst 19.2% (23/120) were layer enterprises, 1.67% (2/120) were pullet enterprises, 8.33% (10/120) were mixed pullet and layer enterprise, 20.0% (24/120) were breeder enterprises, 0.83% (1/120) were turkey enterprise, and 2.5% were duck enterprises. For the purpose of this study a full report of the survey findings has been omitted, however readers are directed to Chapter 3 for a comprehensive description of the survey results.

The contact network between the 120 survey respondents had 16,640 edges representing on- and off-farms movements relating to feed (5,453/16,640; 32.8%), live birds and hatching eggs (7,444/16,640; 44.7%), poultry waste and litter (3,583/16,640; 21.5%) and, poultry products and table eggs (160/16,640; 9.6%). Network statistics have been shown in Table 4.2 but to summarise, there was a large variation in the node degree with some farms reporting only a single movement in comparison to other which had 50 on- and off-farm movements, however despite this range, the majority of farms only reported a small number of movements resulting in a skewed degree distribution. In addition, the average shortest path between any two pair of farms in the network was only 2.42 whilst there was also a high clustering coefficient of 0.824 with geographical clustering and network hubs surrounding urbanised centres such as Auckland, New Plymouth and Christchurch (Figure 4.2).

Table 4.2. Summary network statistics for the New Zealand commercial poultry network showing the yearly on- and off-farms movements relating to feed, live birds, hatching eggs, poultry waste, dead birds, litter, and all other poultry products.

Network statistics	Commercial poultry
<i>Number of nodes</i>	120
<i>Number of links</i>	16640
<i>Density</i>	2.00×10^{-1}
<i>Diameter</i>	7
<i>Mean degree (min-max)</i>	29.00 (1-50)
<i>Mean betweenness (min-max)</i>	81.44 (0-740)
<i>Average path length</i>	2.42
<i>Clustering coefficient</i>	0.824
<i>Total number of SCC^a</i>	NA
<i>(Number of nodes in GSCC^a)</i>	
<i>Total number of WCC^b</i>	3 (188)
<i>(Number of nodes in GWCC^b)</i>	
<i>Reciprocity</i>	NA
<i>Fragmentation</i>	3.32×10^{-2}

^a SCC: Strongly Connected Component - A set of nodes within the network in which all nodes are mutually accessible by following the direction of the links. The giant strongly connected component (GSCC) is the largest SCC.

^b WCC: Weakly Connected Component - A set of nodes within the network in which all nodes are mutually accessible ignoring the direction of the links. The giant weakly connected component (GWCC) is the largest WCC.

NA: network statistics not applicable for an undirected network graph

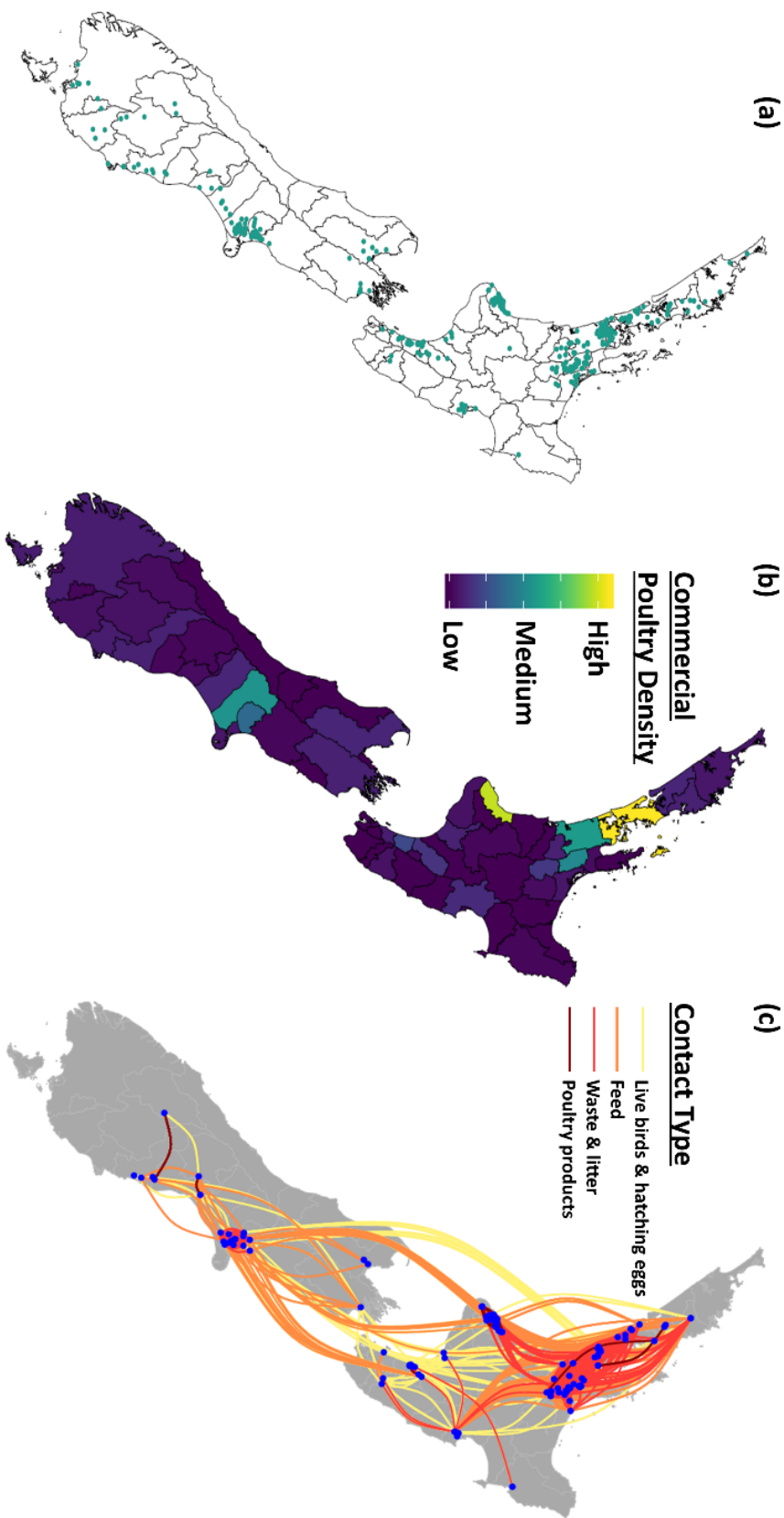


Figure 4.2. (a) Location of commercial poultry enterprises in New Zealand registered with either PIANZ or EPF as of June 2016, (b) choropleth map showing the density of commercial poultry enterprises within the 67 territorial authorities and (c) commercial contact network for the on – and off-farm movement of feed, live birds, hatching eggs, poultry waste, dead birds, litter, and all other poultry products.

4.4.2. Backyard poultry network

4.4.2.1. TradeMe® transactions

Over the seven-year period from 01st January 2012 to 31st December 2018, there were a total of 137,270 recorded poultry trades between 59,225 unique traders registered on the TradeMe® website. A consistent trading pattern was seen across each of the study years with only a small majority of the traders trading more than one type of poultry (*i.e.*, chickens, bantams or ducks). Over the entire study period, only 2.4% (1,456/59,225) of traders traded all three types of poultry compared to 6.8% (4,038/59,225) trading bantams and chickens, 0.2% (136/59,225) trading bantams and ducks, 5.7% (3,364/59,225) trading chickens and ducks, 3.5% (2,126/59,225) trading only bantams, 4.8% (2,856/59,225) trading only ducks, and 76.4% (45,249/59,225) trading only chickens. Descriptive statistics on the frequency of trades, the estimated lower bounds on the number of animals traded, and the number of unique traders stratified by poultry type and year are presented in Tables 4.3, 4.4, and 4.5 respectively. To summarise, the average number of trades made in a year equals 19,610 trades involving 23,768 birds to and from 8,460 active traders. The frequency of poultry trades remained relatively consistent between each year in the study period with the number of trades peaking during in the summer months of October through to February (Figure 4.3). On average, 1.38 birds were recorded under the number of animals traded with a maximum of 70 birds recorded under a single transaction although 93.7% (128,627/137,270) of trades had “1” recorded under the number of animals traded including 89.7% (15,979/17,814) from 01st January 2018 to 31st December 2018.

4.4.2.2. Backyard poultry networks

The trader network constructed from poultry trades between 01st January 2018 and 31st December 2018, had a total of 16,453 unique edges between 13,291 uniquely identified traders whilst the spatial network constructed from poultry trades within the same timeframe had 2,712 unique edges between 134 unique suburbs. Network statistics for

Table 4.3. Frequency of poultry trades occurring through the online auction website TradeMe® in New Zealand stratified by the type of bird (n = 3) and year of trade.

	2012	2013	2014	2015	2016	2017	2018	Total
Poultry								
<i>Chickens</i>	16,895	17,176	16,074	16,632	15,918	15,436	15,113	113,244
<i>Bantams</i>	2,494	2,179	1,909	1,681	1,618	1,494	1,105	12,480
<i>Ducks</i>	1,508	1,598	1,749	1,779	1,715	1,601	1,596	11,546
Total	20,897	20,953	19,732	20,092	19,251	18,531	17,814	137,270

Table 4.4. Estimated lower bounds on the number of poultry traded through the online auction website TradeMe® in New Zealand stratified by the type of bird (n = 3) and year of trade.

	2012	2013	2014	2015	2016	2017	2018	Total
Poultry								
<i>Chickens</i>	18,455	19,157	19,060	21,160	22,956	21,483	21,320	143,591
<i>Bantams</i>	2,574	2,373	2,019	1,821	2,048	1,711	1,252	13,798
<i>Ducks</i>	1,526	1,637	1,900	1,979	1,949	1,867	2,069	12,927
Total	22,555	23,167	22,979	24,960	26,953	23,194	22,572	166,380

Table 4.5. Number of uniquely identified poultry traders utilizing the online auction website TradeMe® in New Zealand stratified the type of bird (n = 3) and year of trade.

	2012	2013	2014	2015	2016	2017	2018	Total
Trader								
<i>Sellers</i>	3,540	3,827	3,785	3,774	3,485	3,289	3,288	14,812
<i>Buyers</i>	11,075	11,625	10,909	11,239	11,091	11,079	10,944	53,268
Total	13,232	14,051	13,371	13,755	13,470	13,422	13,291	59,225



the trader network and spatial network have been summarised in Tables 4.6 and 4.7, respectively, with additional statistics for the subsequent study years presented in Appendix C (Tables C1 and C2). To summarise, all the trader networks for the different poultry type (*i.e.*, chicken, bantams and ducks) in 2018 only had a small proportion of all possible connections actually realised with a density of 9.3×10^{-5} for the network constructed from all poultry trades. Networks varied in their diameter, ranging from 5-24 and their average path length ranging from 1.37-8.55 emphasising the difference between the number of traders and frequency at which chickens are traded in comparison to bantams and ducks.

The reciprocity, fragmentation and clustering coefficients between the networks were reasonable similar. Reciprocity equalled zero for the trader networks showing the transaction of bantams and ducks whilst being only slightly above zero for the chicken trader network. This indicates that only a small fraction of bi-directional links are present in the network. Fragmentation was slightly below one for all the networks whilst the clustering coefficients all ranged between 0.001 and 0.006 indicating an overall lack of cohesiveness between all the traders in the networks. This result is further emphasised by the small number of traders identified in the GSCCs in each graph. For example, when all poultry transaction are considered the network contains 13,200 SCC with the largest GSCC only containing 0.6% (89/13,291) of traders however, when the direction of the trade is ignored the network contains 630 WCC with the largest GSWC containing 89.1% (11,847/13,291) of traders (Table 4.6). This highlights the limited number of bidirectional trades and emphasises the fragmentation in the network.

Table 4.6. Summary network statistics for the trader network showing poultry trades occurring through the online auction website TradeMe® in New Zealand from 01st January 2018 to 31st December 2018.

Network statistic	Chickens	Bantams	Ducks	All Poultry
<i>Number of nodes</i>	11,804	1,087	1,647	13,291
<i>Number of links</i>	14,000	1,002	1,512	16,453
<i>Density</i>	1.01×10^{-4}	8.49×10^{-4}	5.58×10^{-4}	9.32×10^{-5}
<i>Diameter</i>	24	5	5	20
<i>Mean in-degree (min-max)</i>	1.19 (0-42)	0.92 (0-10)	0.92 (0-15)	1.24 (1-50)
<i>Mean out-degree (min-max)</i>	1.19 (0-395)	0.92 (0-63)	0.92 (0-39)	1.24 (0-395)
<i>Mean betweenness (min-max)</i>	364.30 (0-266,038)	1.72 (0-521)	0.47 (0-123)	538.40 (0-382,789)
<i>Average path length</i>	8.55	1.91	1.37	7.56
<i>Clustering coefficient</i>	1.24×10^{-3}	1.57×10^{-3}	5.33×10^{-3}	1.62×10^{-3}
<i>Total number of SCC^a</i>	11,738	1,085	1,647	13,200
<i>(Number of nodes in GSCC^a)</i>	(64)	(3)	(1)	(89)
<i>Total number of WCC^b</i>	587	161	229	630
<i>(Number of nodes in GWCC^b)</i>	(10,428)	(677)	(1,004)	(11,847)
<i>Reciprocity</i>	5.714×10^{-4}	0.00	0.00	8.509×10^{-4}
<i>Fragmentation</i>	1.00	1.00	1.00	1.00

^a SCC: Strongly Connected Component - A set of nodes within the network in which all nodes are mutually accessible by following the direction of the links. The giant strongly connected component (GSCC) is the largest SCC.

^b WCC: Weakly Connected Component - A set of nodes within the network in which all nodes are mutually accessible ignoring the direction of the links. The giant weakly connected component (GWCC) is the largest WCC.

Table 4.7. Summary network statistics for the spatial network showing poultry trades occurring through the online auction website TradeMe® in New Zealand from 01st January 2018 to 31st December 2018.

Network statistic	Chickens	Bantams	Ducks	All Poultry
<i>Number of nodes</i>	132	112	118	133
<i>Number of links</i>	2424	447	636	2712
<i>Density</i>	1.40x10 ⁻¹	3.60x10 ⁻²	4.61x10 ⁻²	1.54x10 ⁻¹
<i>Diameter</i>	6	8	6	4
<i>Mean in-degree (min-max)</i>	1.19 (0-42)	0.92 (0-10)	0.92 (0-15)	1.24 (0-50)
<i>Mean out-degree (min-max)</i>	1.19 (0-395)	0.92 (0-63)	0.92 (0-39)	1.24 (0-395)
<i>Mean betweenness (min-max)</i>	364.30 (0-266038)	1.72 (0-521)	0.47 (0-123)	53.40 (0-382789)
<i>Average path length</i>	2.04	3.06	2.85	2.00
<i>Clustering coefficient</i>	0.480	0.243	0.298	0.504
<i>Total number of SCC^a</i>	27	45	39	23
<i>(Number of nodes in GSCC^a)</i>	(106)	(68)	(80)	(111)
<i>Total number of WCC^b</i>	1	1	1	1
<i>(Number of nodes in GWCC^b)</i>	(132)	(112)	(118)	(133)
<i>Reciprocity</i>	0.446	0.246	0.311	0.476
<i>Fragmentation</i>	0.36	0.63	0.54	0.30

a SCC: Strongly Connected Component - A set of nodes within the network in which all nodes are mutually accessible by following the direction of the links. The giant strongly connected component (GSCC) is the largest SCC.

b WCC: Weakly Connected Component - A set of nodes within the network in which all nodes are mutually accessible ignoring the direction of the links. The giant weakly connected component (GWCC) is the largest WCC.

Within each network, there was a huge range in the node degree although the vast majority of poultry traders only had a degree of one. For example in the trader networks showing transactions for all poultry 69.9% (9,284/13,291) and 14.5% (1,924/13,291) of traders had a degree of one or two, respectively, compared to a smaller proportion with degrees ranging between 3 to 10 (1,732/13,291; 13.0%), 11 to 25 (218/13,291; 1.6%), and 26 to 395 (133/13,291; 1.0%). This skewed degree distribution suggests that a small majority of the nodes are responsible for the majority of connections which is likely to lead to the formation of network hubs. Highly connected individuals acting as hubs in the poultry trader network are highlighted in Figure 4.4. Similar variation in the network

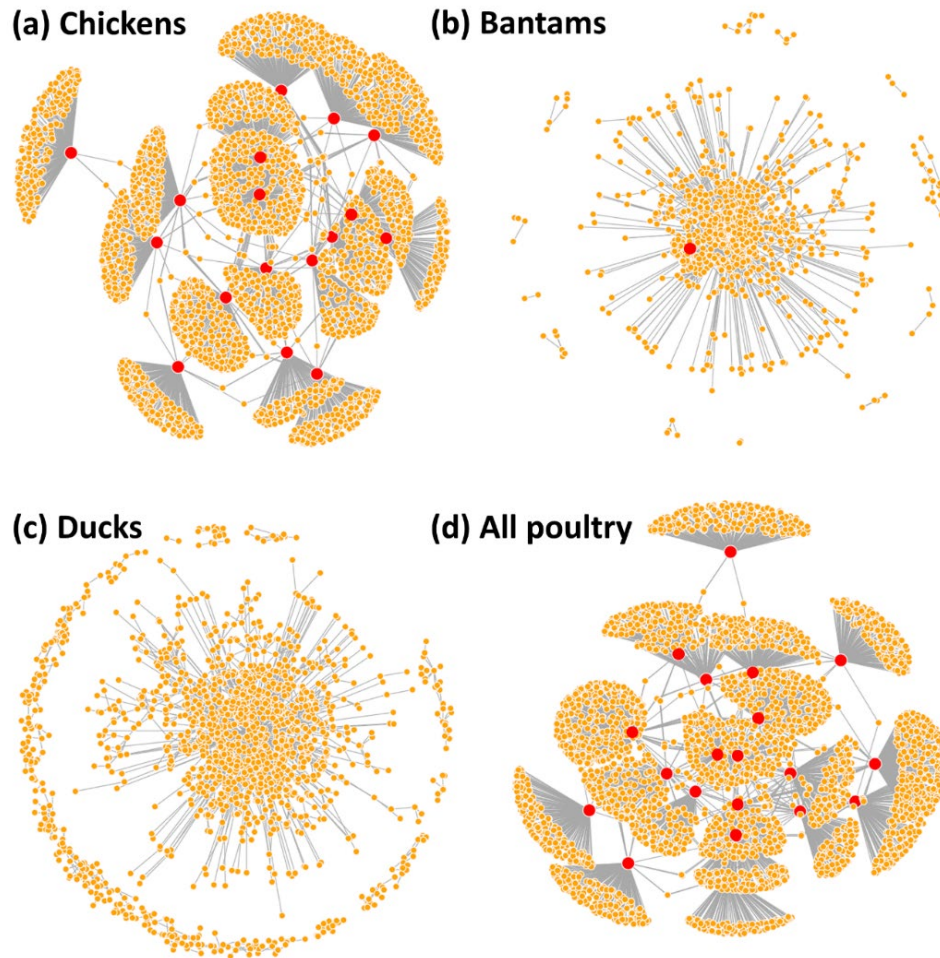


Figure 4.4. Trader contact networks for (a) chicken*, (b) bantam, (c) duck and (d) all poultry* trades occurring through the online auction website TradeMe® in New Zealand from 01st January 2018 to 31st December 2018. Node colour and size highlights those nodes with a degree centrality measure ≥ 50 for bantam and duck trades (red: degree ≥ 50 , orange: degree < 50) and ≥ 100 for chicken and all poultry trades (red: degree ≥ 100 , orange: degree < 100). To improve clarity in the visualization, only connections to and from nodes with a degree ≥ 100 are shown in the networks showing (a) chicken and (d) all poultry trades.

statistics was seen between all spatial networks for the different poultry type however all the spatial networks appeared to have a higher level of cohesiveness. For example, when considering all poultry transactions, the observed network reciprocity was 0.4, fragmentation was 0.3, the clustering coefficient was 0.5 and 82.8% (111/134) of the

suburbs formed the GSCC. Figure 4.5 further highlights the number of connections between all regions of the country with particularly strong linkages between the major urban centres in New Zealand.

4.4.2.3. Backyard poultry disease simulation model

Out of the 10,000 simulation runs, 9,642 (96.4%) resulted in all 134 suburb nodes becoming infected while the remainder were generally outbreaks that were seeded on suburb nodes with only inwards contacts. The time it took for any individual suburb node to become infected during an outbreak ranged between 0 to 2,555 days with a mean equal to 176 days. The full summary data on the number of days until infection for each individual suburb node is presented in Appendix C (Table C3). Across all 10,000 simulations, the number of suburb nodes that became infected during the first 14 days ranged between 1 to 103 suburbs, with a mean equal to 24, whilst the number of suburb nodes that became infected during the first 30 days ranged between 1 to 123 suburbs, with a mean equal to 51. The top 5 suburbs most frequently infected within the first 14 days of a simulation were Manukau City, Auckland City, Waitakere City, Hamilton, and North Shore.

4.4.3. Migrating wilds birds and waterfowl

In total 983,257 reports identifying bird species had been documented through the New Zealand eBird monitoring website from 01st January 2012 to 31st December 2018, overall sighting 304 bird species including migratory and non-migratory birds. The top five regions with the highest number of reported species were Southland (202 species), Far North (186 species), Taupo (185 species), Waitaki (185 species) and Marlborough (163 species) (Figure 4.6). When reports were crossed reference with the list of migratory birds retrieved from the online Encyclopaedia of New Zealand birds 73,990 reports remained sighting 80 different bird species known to migrate to, from or within New Zealand. The majority of these migratory bird species belonged to the order Charadriiformes: a diverse

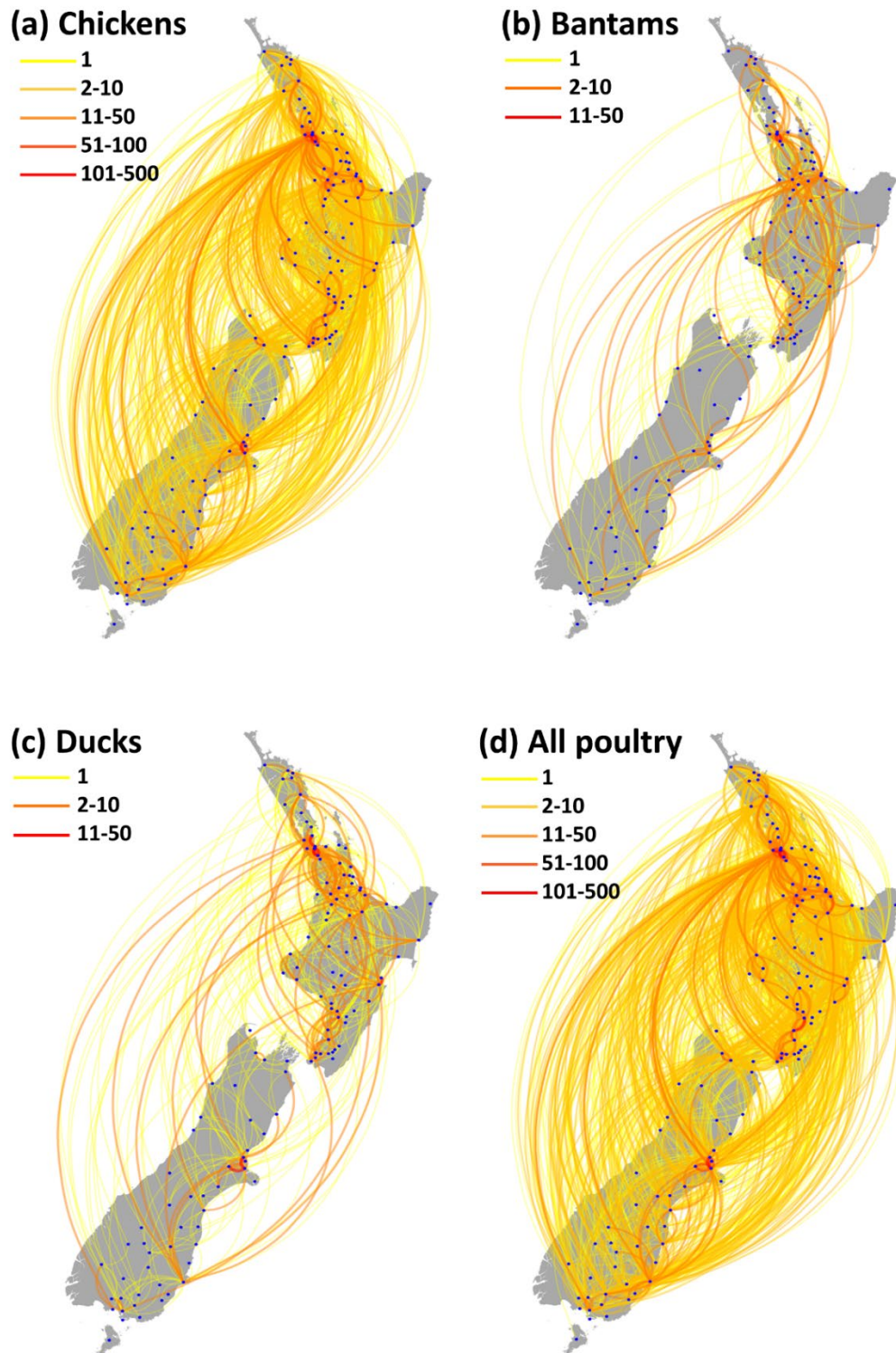


Figure 4.5. Spatial networks for (a) chicken, (b) bantam, (c) duck and (d) all poultry trades occurring through the online auction website TradeMe® in New Zealand from 01st January 2018 to 31st December 2018. Edge colour indicates the frequency of trades (*i.e.*, the number of days between two consecutive trades going to and from the same two nodes).

order of birds containing coastal seabirds and wading birds, however for this analysis, a large proportion (48/80; 60.0%) of the species in this order were excluded from further analysis resulting in a final data extract containing a total of 72,749 reports sighting 32 different bird species (Appendix C, Table C4) from across 10,000 unique locations. Out of these species only three; the South Island oystercatcher (*Haematopus finschi*), the double-banded plover (*Charadrius bicinctus*), and the wrybill (*Anarhynchus frontalis*) were identified as resident birds that migrate only within New Zealand with sightings of these three species accounting for 27.1% (19,692/72,749) of the observations in the extracted data. From the online Encyclopaedia of New Zealand birds an additional 24 bird species belonging to the Anatidae family (Appendix C, Table C5) were identified of which 91.7% (22/24) had been reported to eBird at least once with a total of 90,768 observations.

During the period between 01st January 2018 and 31st December 2018, a total of 29,080 reports had been documented in eBird including 8,109 sightings of resident bird species known to migrate within New Zealand, 20,970 sightings of exotic migratory bird species, and 24,182 sightings of water birds belonging to the Anatidae family. Observations had been made across New Zealand however the top three regions with the highest number of bird reports across all species was Auckland, Hauraki, and Waimakariri with a similar pattern seen when the species are grouped such that Auckland had the highest count of resident bird species, exotic bird species, and water birds. However, after taking into account the regional population density the top three regions were Southland, Mackenzie, and Westland for all bird species known to migrate to, from or within New Zealand, and Southland, Mackenzie and Clutha for water birds (Figure 4.6). A similar pattern can be seen across all the study years although the total number of reports made each year can be seen to dramatically increase (Appendix C, Figures C3-C8).

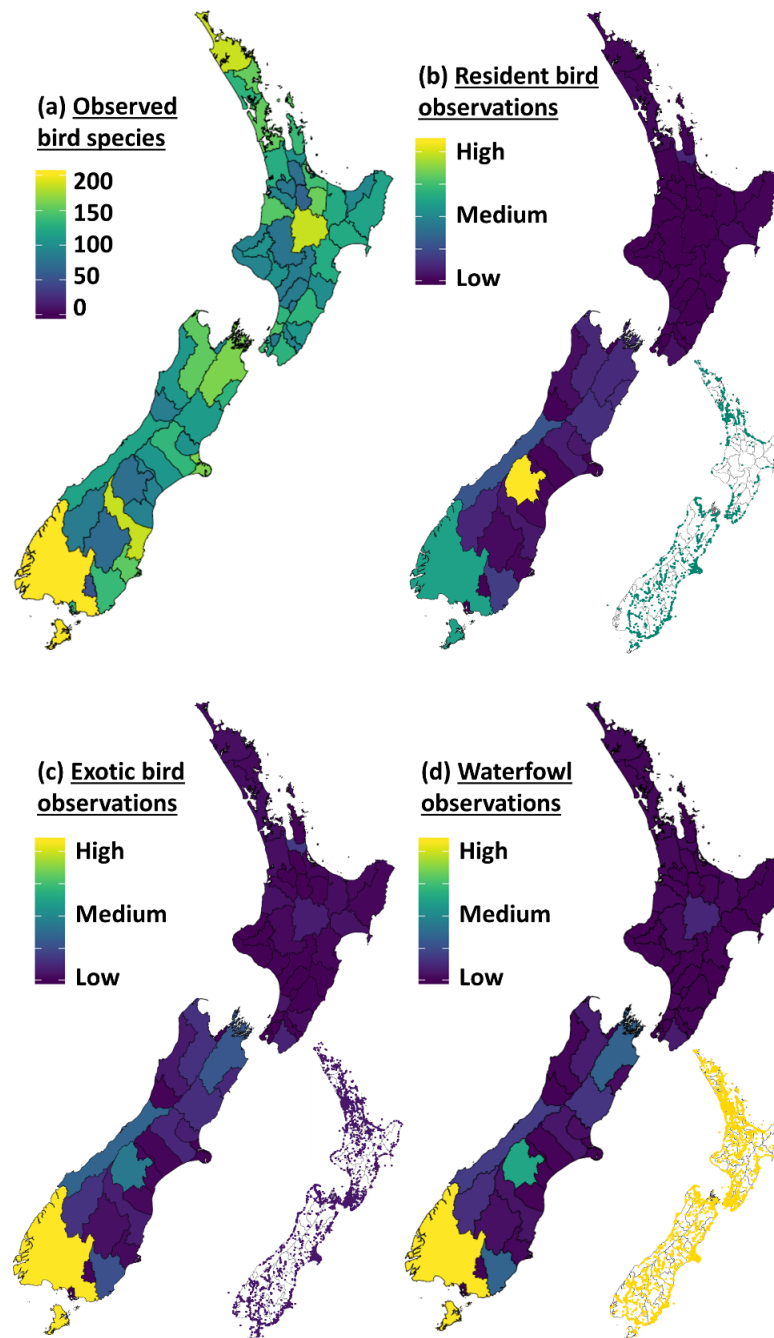


Figure 4.6. Choropleth maps showing the number of bird sighting reported to eBird in each of the 67 territorial authorities from 01st January 2018 to 31st December 2018 with (a) indicating the total number of bird species reported, (b) the number of resident wild birds that are known to migrate within New Zealand, (c) the number of exotic wild birds known to migrate to and from New Zealand, and (d) the number of resident water birds within the Anatidae family. The raw point data is presented alongside each map, with the choropleth maps adjusted for by the population density in each region.

4.4.4. Disease transmission risk between commercial and non-commercial poultry

Overall, the bivariate map showing spatial overlap between commercial poultry premises and backyard poultry trade highlights a number of regions distributed across New Zealand that have a high annual degree (i.e., a high number of movements over a single year), suggesting that there is a lot of backyard poultry moving to, from or within the region and a high density of commercial poultry premises; including the Far North, Kaipara, Waipa, South Waikato, Kapiti Coast and Waitaki (Figure 4.7). In comparison to backyard poultry, the spatial overlap between commercial poultry premises and wild birds, both resident and exotic, is far more limited with many regions with a high number of wild bird observations having very few or no commercial poultry premises present although, out of those limited number of high risk regions most have been identified on the North Island and are also those regions with a high number of backyard poultry trades including; the Far North, Whanagarei, Waikato, Hauraki, Whanganui and Hastings (Figure 4.7).

For backyard poultry, there is a greater number of regions on the South Island; including the Tasman, Dunedin, Timaru, and Waimakariri, that have been identified as having a high risk of contact between backyard poultry and resident wild birds in comparison to the risk of contact with exotic wild bird, for which there is a greater number of high risk regions in the North Island (Figure 4.7). Lastly, the bivariate maps showing spatial overlap between water birds and both commercial poultry and risk regions for contact between backyard poultry and water birds than commercial poultry and water birds including; the Tasman, Dunedin, Waimakariri, Waikato and Auckland, although there are also a number of regions with a high number of water bird observations and a fairly moderate density of commercial poultry enterprises (Figure 4.8). The bivariate maps showing the spatial overlap between water birds and migratory birds, highlight large areas across New Zealand where these populations may come into contacts with many high-risk regions

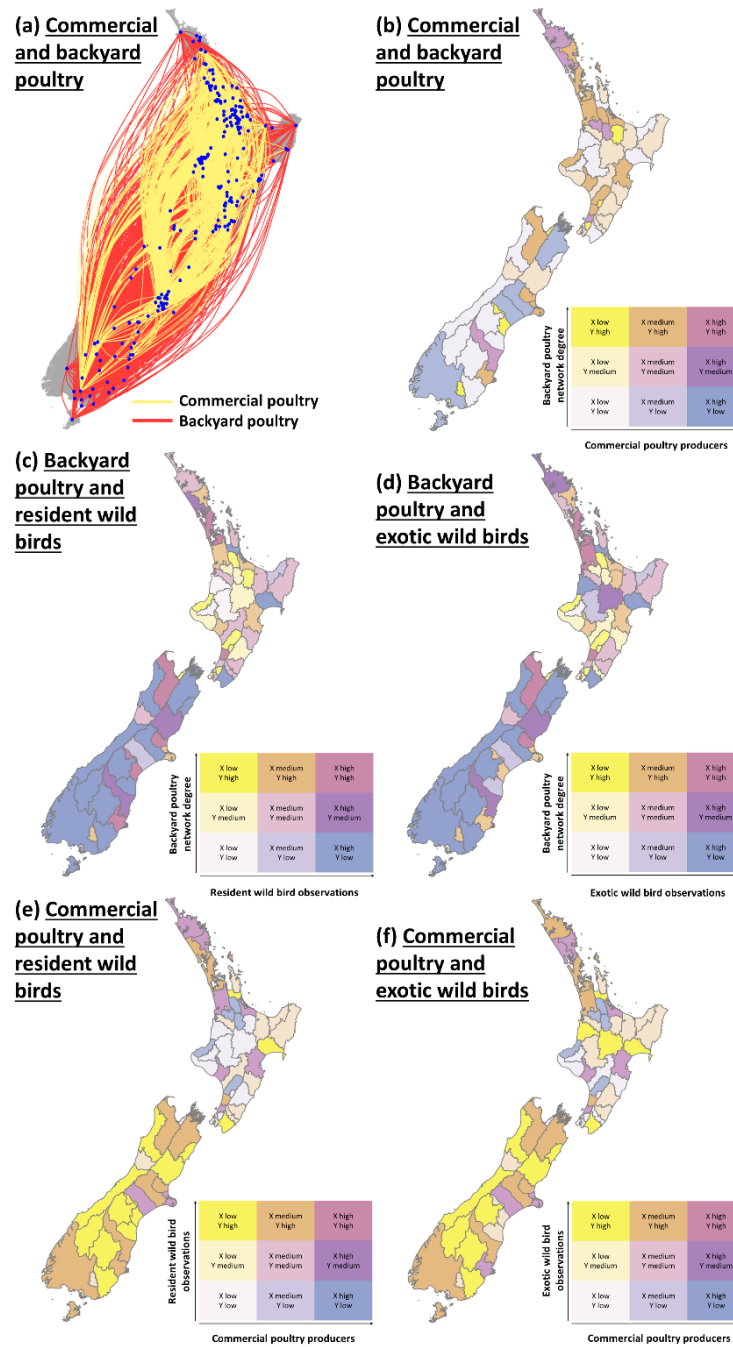


Figure 4.7. The spatial overlap between (a) the commercial poultry network (yellow edges) and the backyard poultry trade network (red edges), and bivariate choropleth maps showing further overlap between (b) commercial poultry enterprise (x-axis) and backyard poultry trade (y-axis), (c) resident migratory wild birds (x-axis) and backyard poultry trade (y-axis), (d) exotic migratory wild birds (x-axis) and backyard poultry trade (y-axis), (e) commercial poultry enterprises (x-axis) and resident migratory wild birds (y-axis), and (f) commercial poultry enterprises (x-axis) and exotic migratory wild birds (y-axis). All data extracted from 01st January 2018 to 31st December 2018.

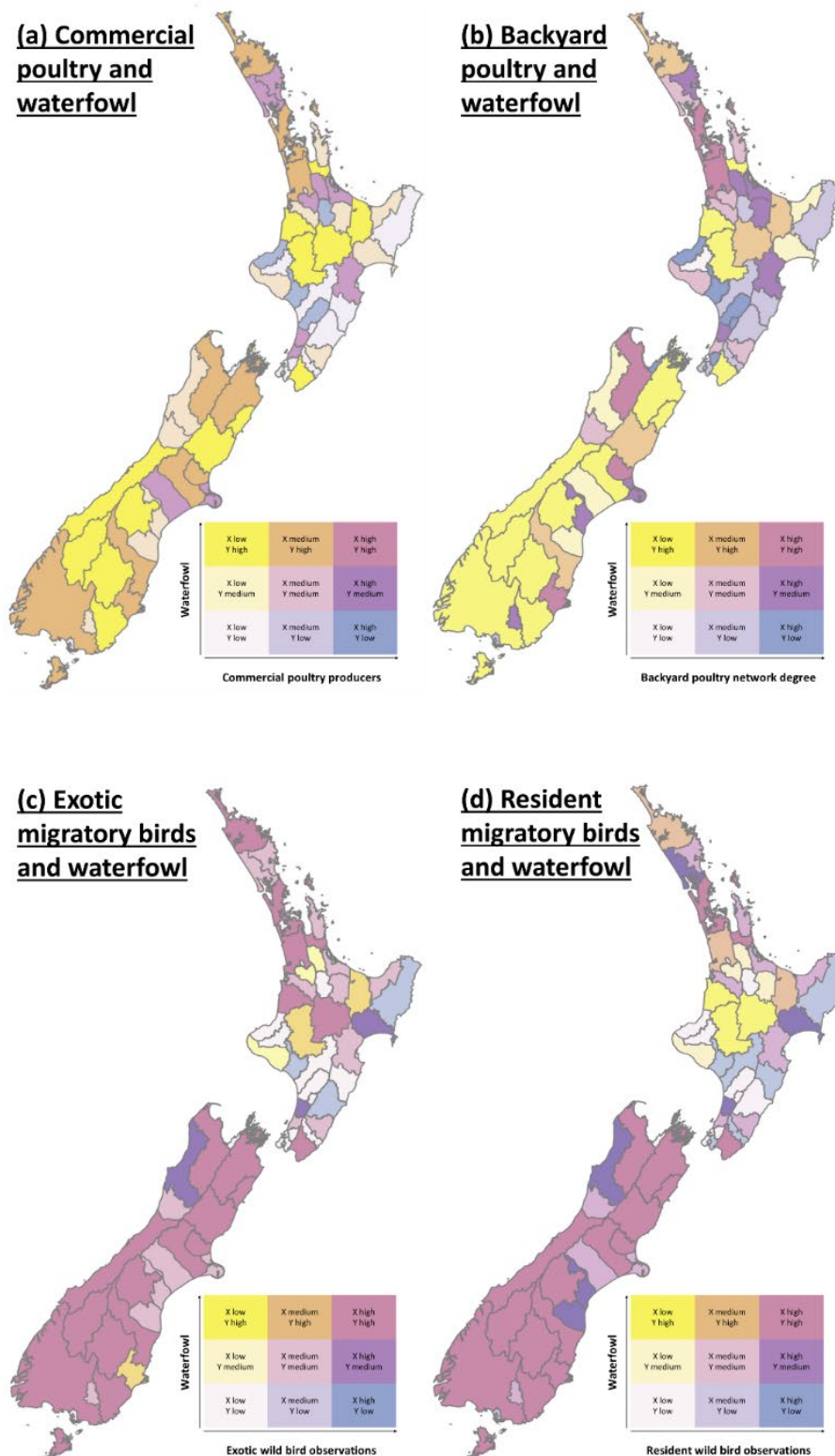


Figure 4.8. Bivariate choropleth maps showing the spatial overlap between resident waterfowl birds in the Anatidae family (y-axis) and (a) commercial poultry enterprises (x-axis), (b) backyard poultry trade (x-axis), (c) exotic migratory wild birds (x-axis), and (d) resident migratory wild birds. All data extracted from 01st January 2018 to 31st December 2018

across the South Island for both exotic and resident migratory bird species and several more regions in the North Island with overlap between water birds and exotic migratory birds.

4.5. Discussion

This study is the first to our knowledge that combines data from commercial, backyard, and wildlife to estimate the risk of disease transmission through avian populations. Analysis of the online auction data in New Zealand revealed highly active backyard poultry trade networks with strong spatial linkages. This spatial pattern is similar to that identified in the New Zealand commercial poultry network with the major population centres acting as network hubs including Auckland, New Plymouth and Christchurch. These hubs, with a high network degree, account for a large proportion of the movements in both the backyard and commercial network, a characteristic of scale-free networks that has been associated with both an increased risk and shorter time to infection in simulated outbreaks (Christley *et al.* 2005; Gates and Woolhouse, 2015), and suggests that these more central regions may be ideal to locate disease surveillance and control strategies.

Within the commercial poultry network the existence of hubs may be explained by the limited number of businesses supplying all the major poultry producers with the majority of these small businesses operating out of the large urban centres where a large proportion of the country's retail and trade is focused, however, the existence of hubs in the trader networks is not so easily explained. One possible reason is that some commercial layer farms will sell spent hens at the end of a laying cycle, which may be distributed to a large number of non-commercial farmers looking to raise backyard poultry. In fact, out of the survey respondents, 69.2% (83/120) of producers indicated selling either point-of-lay or end-of-lay birds, of which a large majority sighted TradeMe® as a means of advertising birds for sale. In addition to TradeMe®, producers also reported other methods including word-of-mouth or local newspapers, with customers picking birds up directly from the

production premises or in public meeting places. If this is the case, the directionality of these trades from commercial to backyard poultry producers poses a greater risk of disease introduction into the backyard poultry network from commercial poultry premises, however, this relationship remains unclear and without knowing the type of contacts being made it is difficult to fully characterise the disease risk to both sectors. Further work is needed to be able to distinguish between the different contact types linking the two networks as the risk of disease via local spreading mechanisms and the spatial proximity of farms in comparison to the risk from direct linkages though the movement of people and vehicles is not only important for different pathogens but also in understanding the best way to mitigate the risk. Despite this uncertainty, these trades still represent a significant potential for contact to occur between backyard and commercial poultry.

In addition to trade between backyard and commercial premises, some backyard traders may also be acting in a commercial fashion either raising specialty poultry breeds for sale or acting as distributors for other poultry sellers. This behaviour is hard to assess with limited information available about the buyers, sellers, or transaction details but, these breeders could also be acting as hubs in the network with a large number of connections. The increasing popularity of lifestyle blocks may also be another reason for the high number of backyard poultry trades near urban centres with a growing demand in regions such as Waikato, Bay of Plenty, Canterbury and Otago (MPI, 2018). The exact number of these lifestyle blocks remains unclear; however, recent estimates suggest there may be more than 140,000 “lifestylers” across New Zealand (MPI, 2018) with the majority located in peri-urban regions where owners have the space to manage a hobby farm or small-holding whilst also having easy access to urban centres. Important characteristics of these properties, such as the number of animals and species present, are often undetermined making it difficult to fully characterise the risk from these properties, however their prime location breaches the gap between the urban-rural divide with the

potential to facilitate many local spreading mechanisms between backyard and commercial poultry even if there is no direct contact through trade. Given this uncertainty, future research should focus on further describing these individuals with a large number of connections and the type of contacts they make as in the event of a disease outbreak, it would be these individuals that pose the greatest risk of disseminating pathogens within the network. Furthermore, with the identification of these key individuals there could be an opportunity for risk-based surveillance focusing on hubs in the trading network; a strategy that has been shown to be highly effective for disease detection given limited resources (Christley *et al.* 2005; Gates and Woolhouse, 2015).

Given the highly active trading community on TradeMe® a more general approach for risk-mitigation could also be taken by using TradeMe® as a platform to communicate important messages to backyard poultry producers in a response to a disease outbreak. Additionally, platforms such as TradeMe®, with a large captured audience, may also act as valuable channels to educate and encourage good ownership practices such as biosecurity and animal health with many backyard producers having been shown to have less knowledge regarding disease management (Smith and Dunipace, 2011; Burns *et al.* 2013) and a preference for more informal networks such as agricultural suppliers, community forums, and neighbours for solving animal health issues (Zheng *et al.* 2010). In particular, increasing biosecurity within the backyard poultry sector could have a significant impact on reducing disease spread with results from the simulation models highlighting the potential spatial spread of disease as a result of backyard poultry trade. It is also important to note from an avian influenza risk perspective that many active traders dealt with both ducks and chickens, which could promote the co-mingling of different circulating avian strains (WHO, 2005). Many of backyard poultry trades also occurred in the spring and summer months which coincides with the time at which susceptible juvenile wild birds begin to migrate between different waterways in New Zealand; a high-risk period for the spread of avian influenza either from direct contact with wild birds or

from the environment (Watts *et al.*2016). To characterise this risk further and provide insights into which region pose the highest risk at which time periods, future analyses should examine the densities of wild birds by migratory seasons.

The analysis of eBird data further highlights the potential risk of disease transmission from migratory birds with different regions posing a high risk for the spread of endemic diseases via resident birds versus the risk of a disease being introduced via exotic migratory birds, within the backyard poultry sector. However, the regional risk profiles for spread between commercial poultry and migratory bird, both resident and exotic, were very similar. However, despite regional differences, the number of high-risk regions remained low when considering the spatial overlap with migratory wild birds and both backyard and commercial poultry. This result may be explained by the high number of majority bird species that can be classified as wader birds or seabirds that rarely venture inland preferring to roost on exposed beaches, headlands or islets (Williams *et al.*2006) where backyard and commercial poultry premises are unlikely to be located. The exception to this is two migratory forest birds both belonging to the Cuculidae family: the shining cuckoo (*Chrysococcyx lucidus*) and long-tailed cuckoo (*Urodynamis taitensis*). Arriving in August and September, these birds migrate to Pacific islands over winter (Higgins and Davies, 1999) and although both favour forest habitats, they have been sighted in urban parks and rural populated regions where they could potentially come into contact with either backyard or commercial poultry.

Of greater concern is the interactions between migratory birds and water birds due to the potential for transmission of novel diseases between these populations followed by the onward transmission to backyard or commercial producers. Interspecies co-occurrence of migrant and resident birds is very common as many species share common feeding grounds, such as lakes, estuaries, and mudflats, or will congregate and intermingle in flocks whilst roosting (Williams *et al.*2006). Resident birds that then go on to migrate

across large distance within New Zealand, have a significantly higher risk of coming into contact with either backyard or commercial poultry. For example, the native South Island pied oystercatcher (*Haematopus finschi*) will breed on the South Island mainly east of the Southern Alps on riverbeds and farmland, high country grasslands, and in coastal areas adjacent to estuaries and lagoons before migrating inland to North Island wintering sites (Sagar, 2013). Some of this spatial overlap between resident birds and both backyard and commercial poultry can be seen in the bivariate maps considering bird species belonging to the Anatidae family with a number of high-risk regions being identified across New Zealand. This pattern may be less of a concern for commercial producers who will often have biosecurity measures in place to prevent birds and other wildlife from entering the production area and coming into contact with poultry. However, it is clear from the survey results that the biosecurity measures are not always effective as 38.3% (46/120) reported the presence of either wild birds or waterfowl within the same area as production, despite also reporting the use of bird-proof housing. The majority of bird species reported were resident birds including sparrows, starlings, pukekos, plovers, swallows, and fantails, which would have been excluded as they are not classed as migratory birds.

The exclusion of these birds presents a major gap in the risk profile of endemic pathogen such as campylobacter and salmonella, both of which are huge concerns for public health (Scott, 2003). In order to fully characterise the risk from wild birds, future analysis should focus on the movement patterns of other resident birds that are known to regularly come into direct contact with commercial poultry or can be found widespread in gardens and public spaces where they present a risk to backyard producers. For example, during the late 1990's an extended outbreak of *Salmonella enterica* serovar Typhimurium definitive type 160 (DT160) was responsible for over 3,000 human salmonellosis cases and thousands of avian deaths (Alley *et al.* 2002). The rapid spread and prolonged period of

this outbreak was thought to be due to ongoing transmission between different host groups particularly wild bird sources (Bloomfield *et al.* 2017).

In addition to this limitation in characterising the spatial risk of wild birds, there are further data constraints in this study that must be considered. In particular citizen science data, such as that obtained from eBird, typically suffers from many drawbacks as it will often be collected in a less rigorous manner in comparison to more traditional scientific research. For this reason, data in this study only included complete checklists to try and ensure a minimum standard of data quality. Nevertheless, there are still many other additional sources of variability and bias inherent to citizen science data that are more difficult to account for including variability in observer skills and effort, geographic coverage and bias in reporting rates between rare and common species (Isaac *et al.* 2014, Mair and Ruete, 2015; Kamp *et al.* 2016). For example, in eBird data observations are likely to be seasonal, although this was not checked in the study extract, and it can be assumed that people are more likely to be outside making observations in the summer months when the weather is better. In previous studies, various statistical modelling approaches have been used to help account for the variability and bias inherent in citizen science data (Hochachka *et al.* 2012; Bird *et al.* 2014; Chen and Gomes, 2018). These approaches may be important in future studies to be able to fully characterise the risk from wild birds, as well as cross-referencing other data sources such as data from the five-minute bird count project promoted by The Department of Conservation (<https://www.doc.govt.nz/our-work/five-minute-bird-counts/>). In addition to integrating different data sources, future studies may also benefit from limiting wild bird observations geographically and include only those reported close to or on commercial poultry premises. For example, in a study by Scott and colleagues (2018), camera traps were used to assess the presence and interactions of different wildlife species with chickens on Australian commercial poultry farms. In this study, they identified that mynah birds, corvids and Columbiformes were the most common birds found on farm,

however the majority of observations showed no direct contact between chickens and wildlife, suggesting the indirect routes of pathogen transfer may be more significant (Scott *et al.* 2018).

In addition to the exclusion of incomplete eBird report, the number of eBird observations were further adjusted for by the regional population density in an effort to account for regions that may have little to no observations due to their remote location. However, this method does not account for differences in the number of observations due to visitors in the region. For example, in New Zealand there are a number of remote regions, particular in the South Island, that have a limited number of permanent residents but are popular destinations for visitors who are often there to enjoy nature and therefore may be contributing more to citizen science projects such as eBird. Future analyses may consider other methods to account for regions that may have less observation due to their location such as environmental niche modelling; a method that attempts to correct variability in species observation and detection by identifying habitat most frequently associated with each species, such as the level of vegetation, and using those to extrapolate predictions to areas that lack observational data. However, it is also worth considering the added value of this method for characterising disease risk with many of these remote regions having little spatial overlap with either backyard or commercial poultry therefore without knowing further details on migration routes of birds species, these low regions would be likely to remain low risk even after data extrapolation.

Further drawbacks with the data used to construct the contact networks for both backyard and commercial poultry must be considered. For example, although the TradeMe[®] dataset listed all the transactions during the study period, the spatial resolution was fairly coarse making it difficult to accurately assess the potential spatial overlap and proximity of backyard premises to commercial poultry producers. However, despite this limitation, it is clear that highly active trading hubs do exist within the network; highlighting the

potential impact of backyard trades in the event of a disease outbreak and the importance of collecting further data on the backyard poultry sector in order to include them in any disease contingency plans. This would include trying to assess what fraction of backyard poultry trade is captured through TradeMe®, without which it would be difficult to determine if there are other informal networks that may be contributing to disease spread such as other trades occurring through local papers and community web forums. The lack of information regarding within suburb contact also limits the ability to model disease dynamics, as seen in our spatial models which undoubtedly overestimate the true rate of spread. The spatial model also suffers from the assumption that all movements carry the same risk with no regards to risk factors such as the number of birds traded. Unfortunately, it is clear from the TradeMe data that the number of animals listed does not always match the number of animals actually being sold as described in the additional comments. To reliably include this risk factor in the spatial model, for example weight each edge by the average number of birds traded, some further work would be needed in which the number of birds sold is extracted from the trader comments. Additional data such as this would not only improve future models used to inform control and surveillance activities, but also reduce the time and resources regulatory authorities would have to invest in performing contact tracing within suburbs in the event of a disease outbreak.

Within the commercial poultry network, survey results provided a huge amount of information regarding on- and off-farm movements such as the movement frequency and the quantity moved. However, the survey's low response rate limits the ability to fully characterise the network due to a large proportion of missing links including potential network connections that are not accounted for by the movement of goods and services such as the movement of personnel to or from other sites or farms with poultry, an important contact that has thought to be responsible for a number of disease outbreaks overseas including Newcastle disease, chronic respiratory disease and infectious

laryngotracheitis (Rawdon *et al.* 2007). Similar to other surveys, it is also likely that there is some degree of non-response bias and reporter bias, that have been discussed in detail in Chapter 3, but overall may result in further missing links, making it difficult to assess if the network structure accurately reflects the contact patterns in the commercial poultry industry.

4.6. Conclusion

Overall, our study findings highlight how the spatial patterns of online poultry trade, migratory birds and resident water birds can affect the risk landscape within the commercial poultry sector. Results highlight the importance of accounting for the growing number of animals traded through online auction websites, over increasingly long distances and shorter timespans, on disease transmission dynamics in New Zealand. Failure to account for the existence of backyard producers and the high frequency of movements between them could lead to gross underestimation of the potential size and spread of infectious disease outbreaks. However, without knowing more details about the demographic characteristics or intended purpose of the animals traded it is also difficult to fully quantify the potential disease transmission risks from backyard poultry. The spatial overlap between wild birds with both backyard and commercial poultry also highlight the importance of accounting for the movement dynamics of a number of bird species, with a low risk of disease introduction into the commercial poultry industry via direct contact with migratory wild birds but a high risk via indirect contact due to the population dynamics of resident birds that not only present a risk of spreading endemic diseases but also a risk of introducing exotic disease due to the co-mingling of species. It is therefore important that future research tries to further characterise the movement of both backyard poultry and wild birds within New Zealand, such that effective risk-based surveillance programmes can be established to maintain freedom from disease in addition to providing data for modelling activities that may be used to inform response actions in the event of a disease incursion.

4.7. Acknowledgements

The authors are grateful to the TradeMe® Corporation for providing access to the anonymised online auction data as well as the numerous contributors to the eBird database and to the eBird team at the Cornell Lab of Ornithology for carefully curating the data and making it publicly available.

4.8. References

- Alexander DJ. An overview of the epidemiology of avian influenza. *Vaccine* 25, 5637–5644, 2007
- Alley MR, Connolly JH, Fenwick SG, Mackereth GF, Leyland MJ, Rogers LE, Haycock M, Nicol C, Reed CEM. An epidemic of salmonellosis caused by *Salmonella Typhimurium* DT160 in wild birds and humans in New Zealand. *N. Z. Vet. J.* 50, 170–176, 2002
- Barabasi AL. Scale-Free Networks: A Decade and Beyond. *Science* 325, 412–413, 2009
- Behravesh CB, Brinson D, Hopkins BA, Gomez TM. Backyard poultry flocks and salmonellosis: a recurring, yet preventable public health challenge. *Clin. Infect. Dis.* 58, 1432–1438, 2014
- Bird TJ, Bates AE, Lefcheck JS, Hill NA, Thomson RJ, Edgar GJ, Stuart-Smith RD, Wotherspoon S, Krkosek M, Stuart-Smith JF, Pecl GT, Barrett N, Frusher S. Statistical solutions for error and bias in global citizen science datasets. *Biol. Conserv.* 173, 144–154, 2014
- Bloomfield SJ, Benschop J, Biggs PJ, Marshall JC, Hayman DTS, Carter PE, Midwinter AC, Mather AE, French NP. Genomic analysis of *Salmonella enterica* serovar Typhimurium DT160 associated with a 14-year outbreak, New Zealand, 1998–2012. *Emerg. Infect. Dis.* 23, 906–913, 2017
- Burns TE, Ribble C, McLaws M, Kelton D, Stephen C. Perspectives of an underrepresented stakeholder group, backyard flock owners, on poultry health and avian influenza control. *J. Risk Res.* 16, 510–518, 2013
- Chen D, Gomes CP. Bias reduction via end-to-end shift learning: application to citizen science. *Assoc. Adv. Artif. Intell.* arXiv:1811.00458, 2018
- Christley RM, Pinchbeck GL, Bowers RG, Clancy D, French NP, Bennett R, Turner J. Infection in social networks: using network analysis to identify high-risk individuals. *Am. J. Epidemiol.* 162, 1024–31, 2005
- Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, Complex Systems*, 2006. URL: <https://igraph.org>
- Daughton AR, Generous N, Priedhorsky R, Deshpande A. An approach to and web-based tool for infectious disease outbreak intervention analysis. *Sci. Rep.* 7, 1–11, 2017

- Derksen T, Lampron R, Hauck R, Pitesky M, Gallardo RA.** Biosecurity assessment and seroprevalence of respiratory diseases in backyard poultry flocks located close to and far from commercial premises. *Avian Dis.* 62, 1–5, 2018
- Dickinson JL, Zuckerberg B, Bonter DN.** Citizen science as an ecological research tool: challenges and benefits. *Annu. Rev. Ecol. Evol. Syst.* 41, 149–172, 2010
- eBird.** *eBird 2018: Year in review.* Cornell Lab of Ornithology, Cornell University, Ithaca, New York, 2018.
- Whitworth D, Newman SH, Mundkur T, Harris P.** *Wild Birds and Avian Influenza.* In: *Wild Birds and Avian Influenza: an introduction to applied field research and disease sampling techniques.* Food and Agriculture Organization (FAO), Rome, Italy, 2007.
- Fiebig L, Smieszek T, Saurina J, Hattendorf J, Zinsstag J.** Contacts between poultry farms, their spatial dimension and their relevance for avian influenza preparedness. *Geospat. Health* 4, 79–95, 2009
- Frazer J, Rawdon T, Stanislawek W.** Avian influenza surveillance programme. *Surveill.* 35, 14–16, 2008
- Frazer J, Rawdon T, Stanislawek W, McFadden A.** Avian influenza surveillance programme. *Surveillance* 36, 17–18, 2009
- Frazer J, Rawdon T, Stanislawek W, McFadden A.** Avian influenza surveillance programme. *Surveillance* 37, 20–22, 2010
- Fruchterman TMJ, Reingold EM.** Graph drawing by force-directed placement. *Software Pract. Exper.* 21, 1129–1164, 1991
- Garner MG, Dubé C, Stevenson MA, Sanson RL, Estrada C, Griffin J.** Evaluating alternative approaches to managing animal disease outbreaks: the role of modelling in policy formulation. *Vet. Ital.* 43, 285–298, 2007
- Gates MC, Woolhouse ME.** Controlling infectious disease through the targeted manipulation of contact network structure. *Epidemics* 12, 11–19, 2015
- Hamilton-West C, Rojas H, Pinto J, Orozco J, Hervé-Claude LP, Urcelay S.** Characterization of backyard poultry production systems and disease risk in the central zone of Chile. *Res. in Vet. Sci.* 93, 121–124, 2012

- Hardstaff JL, Häslar B, Rushton JR.** Livestock trade networks for guiding animal health surveillance. *BMC Vet. Res.* 11, 82, 2015
- Hernandez-Jover M, Schemann K, Toribio JA.** A cross-sectional study on biosecurity practices and communication networks of poultry exhibition in Australia. *Prev. Vet. Med.* 110, 497–509, 2013
- Hernandez-Jover M, Schemann K, Toribio JA.** Evaluating the risk of avian influenza introduction and spread among poultry exhibition flocks in Australia. *Prev. Vet. Med.* 118, 128–141, 2015
- Higgins PJ, Davies J.** *Handbook of Australian, New Zealand and Antarctic birds.* In: Parrots and Dollarbirds, Oxford University Press, Melbourne, Australia, 1999
- Hochachka WM, Fink D, Hutchinson RA, Sheldon D, Wong W-K, Kelling S.** Data-intensive science applied to broad-scale citizen science. *Trends Ecol. Evol.* 27, 130–137, 2012
- Isaac NJB, Strien AJ van, August TA, Zeeuw MP de, Roy DB.** Statistics for citizen science: extracting signals of change from noisy ecological data. *Meth. Ecol. Evol.* 5, 1052–1060, 2014
- Johnson YJ, Colby MM, Tablante NL, Hegngi FN, Salem N, Pope C.** Application of commercial and backyard poultry geographic information system databases for the identification of risk factors for clinical infectious laryngotracheitis in a cluster of cases on the Delmarva Peninsula. *Int. J. Poul. Sc.* 3, 201–205, 2004
- Jutzi S.** *Stop the disease at its source.* FAO, OMS, OIE, and WorldBank Conference on Avian Influenza and Human Pandemic Influenza, Geneva, Switzerland, 2005.
- Kahle D, Wickham, H.** ggmap: Spatial visualization with ggplot2. *The R J.* 5, 144–161, 2013
- Kamp J, Oppel S, Heldbjerg H, Nyegaard T, Donald PF.** Unstructured citizen science data fail to detect long-term population declines of common birds in Denmark. *Divers. Distrib.* 22, 1024–1035, 2016
- Kaplan EH, Craft DL, Wein LM.** Emergency response to a smallpox attack: the case for mass vaccination. *Proc. Natl. Acad. Sci. USA* 99, 10935–10940, 2002
- Keeling MJ, Danon L.** Mathematical modelling of infectious diseases. *Br. Med. Bull.* 92, 33–42, 2009

- Langstaff IG, McKenzie JS, Stanislawek WL, Reed CEM, Poland R, Cork SC.** Surveillance for highly pathogenic avian influenza in migratory shorebirds at the terminus of the East Asian-Australasian Flyway. *N. Z. Vet. J.* 57, 160–165, 2009
- Lockhart CY, Stevenson MA, Rawdon TG.** A cross-sectional study of ownership of backyard poultry in two areas of Palmerston North, New Zealand. *N. Z. Vet. J.* 58, 155–159, 2010a
- Lockhart CY, Stevenson MA, Rawdon TG, Gerber N, French NP.** Patterns of contact within the New Zealand poultry industry. *Prev. Vet. Med.* 95, 258–266, 2010b
- Mair L, Ruete A.** Explaining spatial variation in the recording effort of citizen science data across multiple taxa. *PLoS One* 11, e0147796, 2016
- Monne I, Fusaro A, Nelson MI, Bonfanti L, Mulatti P, Hughes J, Murcia PR, Schivo A, Valastro V, Moreno A, Holmes EC, Cattoli G.** Emergence of a highly pathogenic avian influenza virus from a low-pathogenic progenitor. *J. Virol.* 88, 4375–4388, 2014
- Ministry for Primary Industries (MPI).** *MPI study of New Zealand lifestyle blocks and animal welfare: Summary of findings.* Ministry for Primary Industries, Wellington, New Zealand, Paper No: 2018/03, 2018.
- Pohjola L, Nykäsenoja S, Kivistö R, Soveri T, Huovilainen A, Hänninen ML, Fredriksson-Ahomaa M.** Zoonotic public health hazards in backyard chickens. *Zoonoses Public Health* 63, 420–430, 2016
- Rawdon T, Zheng T, Adlam B, Williman J, Huang S.** Disease risk pathways associated with backyard poultry-keeping in New Zealand. Part 1: Animal health implications. *Surveillance* 38, 12-19, 2011
- R Core Team.** R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2018. <https://www.R-project.org/>
- Reed KD, Meece JK, Henkel JS, Shukla SK.** Birds, migration and emerging zoonoses: west nile virus, lyme disease, influenza a and enteropathogens. *Clin. Med. Res.* 1, 5–12, 2003
- Saatkamp HW, Geers R, Noordhuizen JPTM, Dijkhuizen AA, Huirne RBM, Goedseels V.** National identification and recording systems for contagious animal disease control. *Livest. Prod. Sci.* 43, 253–264, 1995

- Sagar P.** *South Island pied oystercatcher*. New Zealand Birds Online: The Digital Encyclopaedia of New Zealand Birds. Accessed 18 June 2019.
- Schembri N, Hernandez-Jover M, Toribio JA, Holyoake PK.** On-farm characteristics and biosecurity protocols for small-scale swine producers in eastern Australia. *Prev. Vet. Med.* 118, 104–116, 2015
- Scott E.** Food Safety and foodborne disease in the 21st century. *Can. J. Infect. Dis.* 14, 277–280, 2003
- Scott AB, Phalen D, Hernandez-Jover M, Singh M, Groves P, Toribio JLML.** Wildlife presence and interactions with chickens on Australian commercial chicken farms assessed by camera traps. *Avian Dis.* 62, 65–72, 2018
- Short KR, Richard M, Verhagen JH, van Riel D, Schrauwen EJA, van den Brand JMA, Mänz B, Bodewes R, Herfst S.** One health, multiple challenges: The inter-species transmission of influenza A virus. *One Health* 1, 1–13, 2015
- Smith G, Dunipace S.** How backyard poultry flocks influence the effort required to curtail avian influenza epidemics in commercial poultry flocks. *Epidemics* 3, 71–75, 2011
- Sullivan BL, Wood CL, Iliff MJ, Bonney R, Fink D, Kelling S.** eBird: A citizen-based bird observation network in the biological sciences. *Biol. Conserv.* 142, 2282–2292, 2009
- Tana T, Rawdon T, Stanislawek W.** Avian influenza surveillance programme. *Surveillance* 34, 11–13, 2007
- The US Department of Agriculture (USDA).** *Urban Chicken Ownership in Four U.S. Cities*. The US Department of Agriculture, Washington, United States, 2013
- Wang Y, Jiang Z, Jin Z, Tan H, Xu B.** Risk factors for infectious diseases in backyard poultry farms in the Poyang lake area, China. *PLoS One* 8, e67366, 2013
- Watts J, Rawdon T, Stanislawek W, Cobb S, Tana T, Mulqueen K.** Avian influenza: epidemiology and surveillance in New Zealand. *Surveillance* 43, 6–14, 2016
- Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y.** Evolution and ecology of influenza A viruses. *Microbiol. Rev.* 56, 152–179, 1992

World Health Organisation (WHO). *Cumulative number of confirmed human cases for avian influenza A (H5N1) reported to WHO, 2003-2019.* Avian and other zoonotic influenza monthly risk assessment summary. World Health Organisation, Geneva, Switzerland, 2019

World Health Organisation (WHO). Evolution of H5N1 Avian Influenza Viruses in Asia. *Emerg. Infect. Dis.* 11, 1515–1521, 2005

Wickham H. ggplot2: elegant graphics for data analysis. Springer-Verlag, New York, United States, 2016

Williams M, Gummer H, Powlesland R, Robertson H, Taylor G. *Migrations and movements of birds to New Zealand and surrounding seas.* Science & Technical Publishing, Wellington, New Zealand, 2006

Zheng T, Adlam B, Rawdon TG, Stanislawek WL, Cork SC, Hope V, Buddle BM, Grimwood K, Baker MG, O’Keefe JS, Huang QS. A cross-sectional survey of influenza A infection and management practices in small rural backyard poultry flocks in New Zealand. *N. Z. Vet. J.* 58, 74–80, 2010

**Transmission dynamics of *Campylobacter jejuni*
(ST-6964) in New Zealand's commercial poultry
network**

Sabrina S. Greening, Ji Zhang, Anne C. Midwinter, David A. Wilkinson, Ahmed Fayaz,
Deborah A. Williamson, Marti J. Anderson, M. Carolyn Gates, and Nigel P. French

5.1. Abstract

AIMS: Despite intensive surveillance and control activities, campylobacteriosis remains the most frequently reported foodborne illness in New Zealand with the largest proportion of human cases linked to the consumption of poultry meat. As highlighted by the recent emergence and rapid spread of an antimicrobial resistant *Campylobacter jejuni* Sequence Type (ST)-6964 within New Zealand commercial poultry, it is important to understand the relative contribution of different between-flock transmission pathways to guide recommendations around mitigating disease spread. The aims of this study were (i) to describe the genetic population structure of *C. jejuni* ST-6964 within the New Zealand commercial poultry industry and (ii) to investigate the role of multiple contact pathways on the genetic relatedness of these isolates in order to identify the most likely routes of transmission.

MATERIALS AND METHODS: Whole-genome sequencing was performed on 167 *C. jejuni* ST-6964 isolates sampled from across 30 New Zealand commercial poultry enterprises including 26 broiler flocks and 4 breeding flocks. The genetic relatedness between isolates was examined using whole-genome multilocus sequence type (wgMLST) analysis. Multiple pairwise distance matrices were generated using (i) farm coordinates to calculate both the Euclidean distance and road distance between farms and (ii) an industry survey to calculate the shortest path between farms within three potential contact networks constructed from the on- and off-farms movements of either feed, live birds and hatching eggs or poultry waste and litter. Permutational multivariate analysis of variance (PERMANOVA) and distance-based linear models (DistLM) were used to explore the relative importance of geographical distance and network distances as potential determinants of the pairwise genetic relatedness between the *C. jejuni* isolates.

RESULTS: After the removal of recombinant regions from the 167 *C. jejuni* isolates, 230 polymorphic sites were identified including 204 single nucleotide polymorphic loci. Genes from three *C. jejuni* integrated elements; *CJIE1*, *CJIE1v* and *CJIE4* were detected

with differential distributions amongst all the isolates. The plasmid *pTetO* was also found to be present in 49.1% (82/167) of the isolates, with both *CJIEs* and *pTetO* found in samples from across all the major poultry suppliers. The functional genes *glcD* and *amtB* were found to be truncated in 31.1% (52/167) and 32.9% (55/167) of the isolates, respectively, all of which were sampled from farms belonging to only a single poultry supplier. Within the phylogeny reconstructed from core polymorphic loci, two distinct genetic clusters were identified with the majority of isolates sampled from farms belonging to the same parent company grouping within the same genetic cluster, supporting previous evidence that pathogen phylogeny is associated with poultry suppliers. After controlling for this effect, a significant association was found between the pairwise genetic relatedness of the *C. jejuni* isolates and both the road distance and the network distance of transporting feed vehicles.

CONCLUSION: Overall, this study suggests that the transportation of feed within the commercial poultry industry as well as other local contacts between flocks, such as the movements of wildlife and personnel, may play a significant role in the spread of *C. jejuni*. These results could have important implications for the surveillance and control activities within the commercial poultry industry. However, further work is needed to fully characterise the risk of these pathways and to understand how they could be targeted to reduce the spread of *C. jejuni*.

KEYWORDS: *Campylobacter jejuni*, Contact networks, Phylodynamics, PERMANOVA, Correlation matrix

5.2. Introduction

To control the spread of any infectious disease it is essential to have a good understanding of the mechanisms and pathways through which the infectious agent is spreading within the population. However, many pathogens can utilise multiple transmission pathways, and transmission modes often vary among pathogen strains and host populations (Antonovics *et al.* 2017). This creates a major challenge in many infectious disease outbreaks, as without determining the extent to which different transmission pathways are contributing to disease dynamics it is difficult to recommend targeted control strategies that are both timely and cost-effective (Antonovics, 2017; Webster *et al.* 2017). This is true for many pathogens responsible for foodborne illnesses. For example, *Campylobacter*, one of the leading causes of foodborne gastroenteritis worldwide (Kaakoush *et al.* 2015), is known for its complicated dynamics and multiple transmission pathways that make it difficult to model (Koutsoumanis *et al.* 2016) and control despite the implementation of many targeted control strategies (Lin, 2009; Newell *et al.* 2011).

In New Zealand, source attribution models have identified a range of sources responsible for human campylobacteriosis cases, including both environmental and ruminant sources, however by far the largest proportion of human cases have been linked to the consumption of poultry meat (Müllner *et al.* 2009). The identification of this risk factor resulted in the implementation of numerous regulatory and voluntary control strategies along the poultry supply chain (NZFSA, 2008) and in the three years following their implementation (from 2005-2008), there was a 50% reduction in human campylobacteriosis notifications (Sears *et al.* 2011). However, despite continued control efforts, campylobacteriosis continues to be the most frequently reported foodborne illness in New Zealand, with rates as much as ten times that of the United States and double that of other industrialised countries such as the United Kingdom (Olson *et al.* 2008). Furthermore, in 2014, routine sampling at a sentinel surveillance site in the Manawatu region of New Zealand detected a new and emerging strain of *C. jejuni*

Sequence Type (ST) 6964 (Muellner *et al.* 2016). Subsequent sampling throughout the poultry industry detected this emerging strain across all four major poultry suppliers, responsible for servicing over 90% of the industry (Stafford, 2017), suggesting a major epidemiological shift with transmission occurring between suppliers despite the vertically integrated structure of the New Zealand poultry industry and minimal contact between each supplier. This structure was thought to be one of the main factors contributing to the past association between different *Campylobacter* STs and individual poultry suppliers (Müllner *et al.* 2010); however, given the rapid spread of ST-6964, it is clear that there is a gap in biosecurity that needs to be targeted in order to prevent further spread. Nevertheless, without knowing how different transmission pathways contributed towards the spread of the pathogen, it is difficult to know where to direct control and surveillance activities.

In addition to the change in transmission dynamics, initial evidence from human isolates showed that ST-6964 was also resistant to a range of antimicrobial agents, including tetracyclines and fluoroquinolones. As a result, several surveys were launched sampling both humans and poultry with initial survey results detecting fluoroquinolone-resistance in 30% of the human *C. jejuni* isolates sampled, 77% of which were also resistant to tetracycline, and both ciprofloxacin- and tetracycline-resistance in 37% of *C. jejuni* isolates sampled from poultry (Muellner *et al.* 2016). These findings represented a significant increase in the patterns of resistance (Williamson *et al.* 2015), similar to that seen in many other countries (Kaakoush *et al.* 2015), incentivising further investigations using molecular sequencing data to try and understand the evolutionary mechanisms driving this resistance. The genomic analysis provided evidence of ST-6964 undergoing rapid evolution in New Zealand through multiple mechanisms including the integration of both the *tetO* gene and prophage integrated elements amongst tetracycline-resistant isolates (French *et al.* 2019).

Having this knowledge is important to help characterise the genetic diversity between epidemiologically related isolates (Duong *et al.* 2009) which can then be used to help resolve unknown transmission dynamics (Grenfell *et al.* 2004; Ypma *et al.* 2012; Gilbertson *et al.* 2018) using approaches based on principles in coalescent theory (Kingman, 2000), which, simply put, suggests that as pathogen isolates from infected individuals become more genetically diverse, then the hosts are less likely to be directly linked to each other in the chain of transmission. In addition to many techniques that rely on pathogen sequence data alone to infer transmission dynamics, an increasing number of studies are integrating pathogen phylogenies with host contact network data to help determine which transmission modes are most important for the spread of pathogens within a population (Leventhal *et al.* 2012; Stadler and Bonhoeffer, 2013; Jombart *et al.* 2014).

In the study of infectious disease, contact data tries to capture interactions which may be contributing to transmission between hosts, including both direct and indirect contacts (Sah *et al.* 2018; Silk *et al.* 2018; Chaters *et al.* 2019). Data can be used to reconstruct network graphs and parameterise disease transmission models in order to study the influence of population structure on disease spread and test disease control measures (Christley *et al.* 2005; Bajardi *et al.* 2012; Gates *et al.* 2015). Alongside traditional contact tracing methods contact networks can also be used in disease outbreaks to help reconstruct transmission trees under the assumption that the transmission network will always be a subset of the contact network (Craft, 2015). This largely relies on the ability to correctly define all the contacts that are relevant for disease transmission; however, the challenges in collecting network data are numerous. For example, information on the disease dynamics may be limited, making it difficult to identify all the host species and define the contacts between them. Due to these limitations, many previous studies have relied on network-based disease simulation models (Eubank *et al.* 2004; Zhang *et al.* 2012; Enns and Brandeau, 2011) with few real-world examples currently in the literature to

validate the methodologies used to integrate pathogen sequence data into network analyses.

In New Zealand, the rapid spread of the recently emerged *C. jejuni* ST-6964 presents a timely opportunity to collect pathogen whole-genome sequence data and validate approaches integrating pathogen phylogenies and contact network data with hopes of identifying the relative contribution of different transmission pathways towards the spread of *C. jejuni* ST-6964 throughout the poultry industry. Given this, the aims of this study were (i) to describe the genetic population structure of *C. jejuni* ST-6964 within the New Zealand commercial poultry industry and (ii) to investigate the role of multiple contact pathways on the genetic relatedness of these isolates in order to determine the most important routes of transmission contributing towards the rapid spread of this strain.

5.3. Materials and Methods

5.3.1. Sample collection, isolate culture, and whole-genome sequencing

Swabs from the pooled caecal contents of up to five chickens from the same shed were collected from slaughter processing plants across New Zealand's North and South Island ensuring there were samples taken from each of the four major poultry suppliers (hereafter anonymously referred to as "A", "B", "C" and "D"). Sampling took place between May 2015 and July 2016 as an extension of the poultry survey reported by the Institute of Environmental Science and Research (ESR) in collaboration with the Poultry Industry of New Zealand (PIANZ) (Muellner *et al.* 2016). As samples were taken from poultry carcasses post-slaughter at commercial poultry abattoirs it was advised by the Massey University Human Ethics Committee: Northern that animal ethical approval was not required. Swabs were delivered for bacterial culture to the Molecular Epidemiology and Public Health laboratory (^mEpiLab) at Massey University where cultures were grown in a microaerobic incubator (Don Whitley Scientific, Yorkshire, UK) at 42 °C on modified charcoal cefoperazone deoxycholate agar (mCCDA) (LabM, Lancashire, UK) containing

ciprofloxacin (4 mg/litre) (Sigma, Missouri, USA) and tetracycline (16 mg/litre) (Sigma, Missouri, USA) for selective isolation of resistant *C. jejuni* colonies.

One to two single colonies from each positive plate were sub-cultured and genomic DNA was isolated on a JANUS automated workstation (PerkinElmer, <https://www.perkinelmer.com>) by using Chemagic magnetic bead technology, according to the manufacturer's instructions. The DNA quality was assessed using Qubit™ dsDNA high sensitivity assay kits (Thermo Fisher Scientific Inc.) before DNA libraries were prepared using a NexteraXT DNA preparation kit (Illumina, <https://www.illumina.com>). The majority of the isolates were sequenced using the NextSeq 500 platform (Illumina) *i.e.*, 2 × 100 bp sequencing, as previously described (Baines *et al.* 2016), whilst four representative *C. jejuni* isolates also underwent whole-genome sequencing on the Pacific Biosciences, Inc., RS II platform (<https://www.pacb.com>). The additional DNA extractions protocol and library preparation for these four isolates have been described elsewhere (French *et al.* 2019) and sequence data are available from GenBank BioProject ID PRJNA520992 and PubMLST (<https://pubmlst.org/campylobacter> External Link) nos. 70207–12, 70229, 70230, 70232, 70233, 70252, 70253, and 78631–845.

5.3.2. Genome assembly and whole-genome MLST analysis

The Illumina raw reads underwent a quality control check using the QCtool pipeline developed by Mauro Truglio (<https://github.com/mtruglio/QCtool>) to ensure potential contaminants such as PhiX control reads and adapter sequences were removed. The quality control report was then checked manually to exclude further poor-quality sequencing results which might have been deleterious in the subsequent analysis. Genome *de novo* assembly was completed using the software tool Spades (v3.10.1) (Bankevich *et al.* 2012) with the assembly run in “careful mode” to ensure mismatches were corrected. Assemblies were then imported into Prokka (v1.12) (Seemann, 2014) such that genomic features including coding sequences and ribosomal ribonucleic acid

(rRNA) genes could be annotated after which, the relationship between the isolates was determined by an *ad hoc* wgMLST analysis using Genome Profiler (GeP) (Zhang *et al.* 2015) to convert assembly data into wgMLST allelic profiles.

For this analysis, a reference sequence was selected from the list of assemblies based on the quality control checks ensuring that the reference sequence had the smallest number overlapping reads (*i.e.*, contigs) whilst not having extreme values for maximum length and GC content in comparison to all other isolates. Next, recombinant regions were removed using Gubbins (v2.2.0) (Croucher *et al.* 2015) in order to try and mitigate the effects of horizontal sequence transfer mechanisms on phylogenetic reconstructions (Boto, 2010). After the removal of recombinant regions, the alignment of core polymorphic loci was used to construct a rooted maximum-likelihood phylogenetic tree using the R package *phangorn* (Schliep, 2011), allowing the genetic relatedness of isolates to be visualised. To examine if closely related isolates shared other common factors, metric multidimensional scaling (*mMDS*) ordination plots (Kruskal and Wish, 1978) were generated in PRIMER v7.0 (Clarke and Gorley, 2015) with plots mapping isolates in a two-dimensional Euclidean space in a manner that preserves the dissimilarity scales present in the underlying genetic distance matrices described in section 5.3.4. Different colours were then used to easily identify isolates sharing a common factor such as parent company or network communities described in section 5.3.5.

5.3.3. Comparative genomics of mobile elements

For the purpose of this study, the presence of three plasmids (*pTet*, *pVir* and CP006703), seven *C. jejuni* integrated elements (*CJIE1-6* and *CJIE1v*), and two functional genes (*amtB* and *glcD*) was examined. These were selected on the basis that they have either been previously identified in ST-6964 isolates (French *et al.* 2019) or are known to contribute to the resistance, virulence or host adaptation of *C. jejuni*. For example, the virulence plasmid *pVir* has been associated with *C. jejuni* virulence in a number of in vitro cell

culture systems, animal models and clinical infection (Bacon *et al.* 2002; Tracz *et al.* 2005; Wu *et al.* 2013); and although the mechanisms and role of *pVir* in pathogenesis are not well understood, DNA sequencing and mutational analysis means it has been well characterised (Bacon *et al.* 2002; Shen *et al.* 2016).

To determine the likelihood of the plasmids and *CJIEs* being in each of the *C. jejuni* ST-6964 genomes, a method previously reported by Llarena and colleagues (2016) was used. To summarise, this method uses all the amino acid sequences within the coding region of pre-selected reference sequences (Table 5.1) for each of the mobile elements and run queries that aim to match these sequences with those found in the *C.jejuni* ST-6964 genomes using the online BLASTX tool. The hit with the highest BLAST score in each BLAST search was used to calculate the percentage of the query amino acid sequence covered by the BLAST alignment whilst the mean of each *CJIEs* in each whole-genome sequence was also calculated. If the identity percentage was no less than 70% and the *e*-value; that is, the number of hits expected by chance, no less than 0.0001, the *CJIE* was considered present. This analysis was completed by a custom Perl script (Llarena *et al.* 2016) with results displayed as heat maps alongside the maximum-likelihood phylogenetic tree described above using the online tool Interactive Tree of Life (iTOL) (v4.5.3) (Letunic and Bork, 2016). In addition to plasmids and *CJIEs*, the selection for the truncation of two functional genes *amtB*, linked to ammonium transportation, and *glcD*, a putative glycolate oxidase subunit, was also displayed in the heat maps. The truncation of these genes has previously been associated with a single poultry supplier in New Zealand (Zhang *et al.* 2017) and it is thought that the relatively low number of human cases associated with this poultry supplier may be linked to the observed truncation of these genes.

Table 5.1. The *C. jejuni* reference sequences and the corresponding NCBI references used to determine the likelihood of three plasmids: *pTet*, *pVir* and CP006703, and seven *C. jejuni* integrated elements (*CJIE1-4* and *CJIE1v*) being located in the *C. jejuni* ST-6964 genomes reported in this study (n = 167), using a method previously described by Llarena *et al.* 2016.

Mobile element	<i>C. jejuni</i> reference strain	NCBI reference sequence
<i>pTet</i>	81-176	AY394561
<i>pVir</i>	81-176	AF226280
CP006703	15-537360	CP006703
<i>CJIE1-4</i>	RM1221	NC_003912
<i>CJIE1v</i>	15AR0984	CP035892

5.3.4. Genetic distance matrices

To represent the relationship between each pair of *C. jejuni* ST-6964 isolates, two different types of genetic distance matrices were calculated based on (i) the uncorrected *p*-distance measure, and (ii) an allelic distance measure. The *p*-distance; that is, the proportion of nucleotide sites that differ between a pair of isolates, was calculated using MEGA7 (v7.0.26) (Kumar *et al.* 2016) without making any corrections for multiple substitutions at the same site or differences in the evolutionary rates among sites. The program GeP (v2.2) (Zhang *et al.* 2015) was used to calculate allelic distances as the proportion of alleles that differed between a pair of isolates out of the total number of alleles observed across that pair. Results obtained using these two matrices were highly consistent, so we report here only the results obtained using the allelic distance matrix with results from the *p*-distance matrix provided in Appendix D.

5.3.5. Model matrices

An additional five model matrices were constructed with the aim of relating these, both individually and collectively, to the genetic distances described above. Model matrices included: (i) the geographical Euclidean distance (based on latitude and longitude) between each pair of farms from which isolates were collected from, (ii) the road distance

between farms, and the network distance between farms based on contact networks constructed from the on- and off-farm movements of (iii) feed, (iv) live birds and hatching eggs, and (v) waste and litter.

The geographical distances between farms were calculated by obtaining a postal address for each farm from a database of commercial poultry producers registered with either the Poultry Industry Association of New Zealand (PIANZ) or the Egg Producers Federation of New Zealand (EPF). The addresses provided in the database were checked using Google Maps (2017) to make sure they specified a poultry production site (indicated by the presence of poultry sheds) and not the producers' residential address. Co-ordinates were collected and the two pairwise distance matrices (Euclidean distances and road distances) were calculated using the R packages *geosphere* (Hijmans, 2019) and *gmapsdistance* (Melo et al, 2018), respectively. The matrices were then expanded to express the geographical distance between each pair of isolates with a value of zero in the matrix indicating that the isolates were sampled from the same farm. The matrices were used to create hierarchical dendrograms representing the geographical distance between sequenced isolates using the R package *ape* (Paradis and Schliep, 2019) which were used to construct the tanglegrams described in section 5.3.6.

The network distances between farms was calculated by constructing several contact networks from the reported on- and off-farm movements relating to either (i) feed, (ii) live birds and hatching eggs, or (iii) waste and litter. In each network, nodes represented the farms from which isolates were sampled, with an undirected edge linking nodes (*i.e.*, farms) utilising the same transport company (Figure 5.1). The reported movements were obtained from the results of an industry survey administered to all active poultry producers in New Zealand registered with either PIANZ or EPF as of June 2016. The survey was based on a previous questionnaire conducted in 2006 (Lockhart *et al.* 2010) and modified in collaboration with PIANZ, EPF and the Ministry for Primary Industries

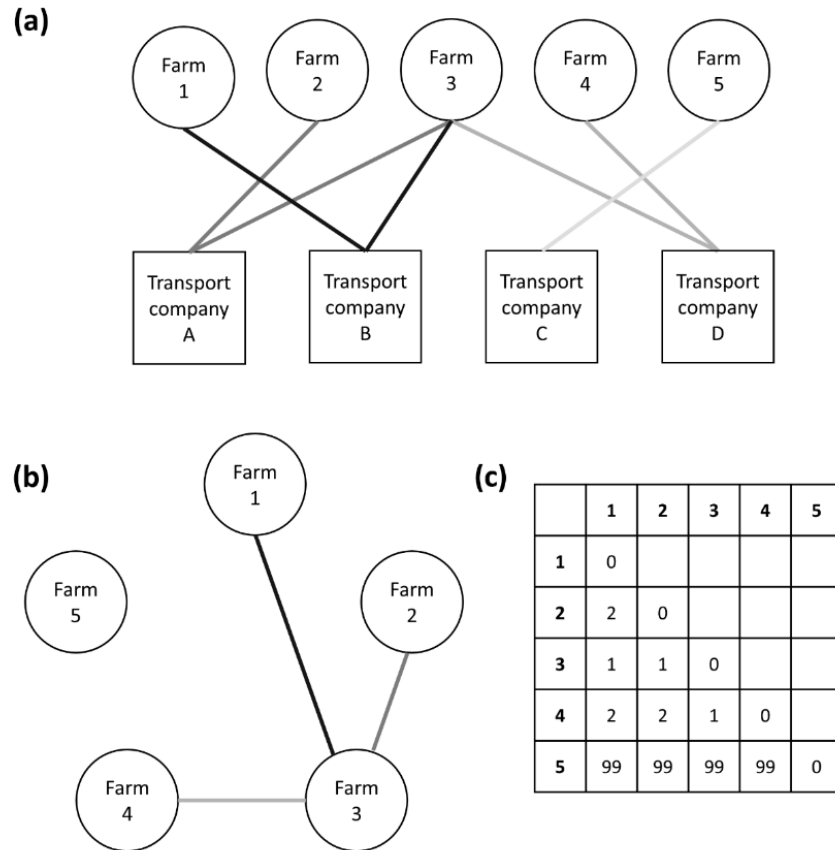


Figure 5.1. Schematic diagram showing the construction of (a) bimodal and (b) unimodal network graphs. In network (a) there are two node types representing either farms (circles) or transporting companies (squares) moving on- and off-farm. In network (b) the farms have been directly linked if they have a transporting company in common. The shortest path (SP) between any two farms is the least number of steps needed to get to one farm from another in the unimodal network graph. For example, in network (b) Farm 4 can be reached from Farm 2 via Farm 3 with a SP equal to 2 as shown in (c) which is the shortest path matrix when considering all the transporting companies with 1 indicating there is a direct link between two farms in (b) and 99 indicating no link.

(MPI) with the aim of collecting information on the farm demographics, contact patterns, and biosecurity practices of New Zealand commercial poultry operations. The study was judged to be low risk thorough peer evaluation and consequently was not formally reviewed by any of the University's Human Ethics Committees. Full details on the survey design and implementation have been described in Chapter 3 and a copy of the complete

survey questionnaire is provided in Appendix A. As not all farms from which isolates were collected completed the survey (*i.e.*, there were non-responders), two networks were constructed for each type of on- and off-farm movement: one including only the farms that responded (hereafter referred to as the “empirical network”), and one including all the farms with sampled isolates (hereafter referred to as the “imputed network”). For the latter, missing links caused by non-responders were inferred based on information supplied by an industry representative who was able to provide the names of the transporting companies that each non-responder was most likely utilising, based on their expert opinion. Results obtained using empirical versus imputed networks were highly congruent, so we report here only the results obtained using the imputed networks with results from the empirical networks provided in Appendix D.

Network graphs were constructed using the R package *igraph* (Csardi and Nepusz, 2006) based on a force-based algorithm proposed by Fruchterman and Reingold (1991). We report here the network degree centrality and betweenness centrality for each network but note that other network statistics for these data have been described in the previous chapters. Network graphs were used to produce pairwise distance matrices with values representing the shortest path (*i.e.*, the minimum number of links) between each pair of farms in each network (Figure 5.1). If a pair of farms were completely unconnected in the network, a numeric value of 99 was recorded in the matrix to represent a very large distance between pairs of isolates that were unconnected. In addition to network matrices, a community analysis was performed in which all the on- and off-farm movements were combined to produce a single network. Communities could then be identified with farms in the same community having more internal links between them than external links to other communities within the network. This analysis was completed using a link community detection algorithm (Ahn *et al.* 2010) in the R package *linkcomm* (Kalinka and Tomancak, 2011).

5.3.6. Relating genetic distances and model matrices

The relationship between the genetic distance matrix and each of the individual model distance matrices was examined using a non-parametric Mantel test (Mantel, 1967). A robust version of the Mantel test was implemented in the RELATE routine in PRIMER (v7.0; Clarke & Gorley 2015), using Spearman's rank correlation (ρ) as a measure of matrix correlation. Model matrices that had a statistically significant relationship with the genetic distance matrix ($p < 0.05$, 9999 permutations) were explored further in the formal linear models described in section 5.3.7. To further examine if an association remained after removing any potential effects of individual farms and their parent company on the genetic distances, a residual genetic distance matrix (removing the effects of parent company and farms nested within parent company) was obtained using the method described in Anderson (2017). Mantel tests were then repeated (using the RELATE routine) to examine the correlation between this residual genetic distance matrix and each of the model matrices. To visualise the relative strengths of matrix associations, a second stage *m*MDS ordination plot (Somerfield and Clarke, 1995) was created after calculating the matrix correlations between all pairs of distance matrices (*i.e.*, the allelic distance matrix, the geographic distance matrices and the network distance matrices). This allowed us to see not only the proximity of each model matrix to the genetic distance matrix but also showed visually the similarities among the various different model matrices.

In addition to the Mantel test and *m*MDS ordination plots, the relationship between pathogen phylogeny and the geographical proximity between all the farms (*i.e.*, Euclidean distance and road distance) was visualised using tanglegrams. Tanglegram were constructed using the R package *dendextend* (Galili, 2015) such that, isolates on the maximum-likelihood phylogenetic tree were connected via auxiliary lines to the corresponding isolate on the hierarchical dendrograms described above representing either the Euclidean distance or road distance between the isolates. To optimise each

tanglegram, a two-tree crossing minimization technique was used to minimize the number of crossings between the auxiliary lines. Tanglegrams were annotated to indicate the production type and parent company of the farm from which the isolates were sampled, and a Spearman's rank correlation coefficient calculated between the trees cophenetic distances matrices (Sokal and Rohlf, 1962).

5.3.7. Explaining variation in genetic distances

To find parsimonious models to explain variation in genetic distances using the information provided in the geographical and/or network distance matrices, we first generated Euclidean coordinates that would capture the information contained in each model distance matrix. This was done by calculating a number (m) of coordinate axes using m MDS to represent the information in each distance matrix. We found that, in every case, $m = 2$ or 3 was sufficient to capture the salient information contained in each of the model distance matrices (stress < 0.01). We also created regression coordinates *i.e.*, in the form of analysis of variance (ANOVA) contrasts, that coded for two additional factors; parent company and farms (nested in parent company) in order to fit the matrix models and ANOVA factors in a regression setting. Note that coordinates corresponding to specific network distance matrices (or ANOVA factors) were kept together as a group for model selection. A DISTLM routine was then used in the PERMANOVA+ add-on package (Anderson *et al.* 2008) for PRIMER v7.0 (Clarke & Gorley 2015) to fit distance-based redundancy analysis (dbRDA) (Legendre & Anderson 1999; McArdle & Anderson 2001) including models of genetic distances versus the network models and ANOVA factors.

To select the final model, we first performed marginal tests for each set individually (with p -values obtained using 9999 unconstrained permutations), and used forward selection based on R^2 to uncover potential redundancies and regions of overlap in the explanatory power of regression variable sets (with p -values for sequential conditional tests at each

step obtained using permutation of residuals under a reduced model) (Freedman and Lane 1983). Next, we then searched for the best overall model; that is combinations of sets of regression coordinates, to explain variation in the genetic distance among isolates using the multivariate analogue to the Akaike Information Criterion (AIC) criterion (Anderson *et al.* 2008) in the DISTLM routine with each model forcing the inclusion of parent company and farm (ANOVA terms in the model) during model selection.

5.4. Results

5.4.1. Sample collection, isolate culture, and whole-genome sequencing

Overall, swabs were taken from 922 birds originating from across 75 commercial poultry farms including farms belonging to all four of the major poultry suppliers. From these, 668 samples from across 41 of the farms gave growth that resembled *Campylobacter* on mCCDA plates containing both ciprofloxacin and tetracycline. Currently only approximately a third of these samples have been sequenced all of which have been confirmed as *C. jejuni* ST-6964, with the ST of the remaining samples unconfirmed. For this study, a subset containing 167 of the confirmed *C. jejuni* ST6964 samples were randomly selected for whole-genome sequencing resulting in isolates from across 30 individual poultry farms belonging to three out of the four major poultry suppliers (hereafter anonymously referred to as “A”, “B” and “C”), and including 26 broiler flocks (*i.e.*, poultry growers) and 4 breeding flocks. The geographical distribution amongst the selected subset of farms was limited to three regions on the North Island of New Zealand with four farms from both suppliers A and B located in the Auckland region, two farms from supplier B and eleven farms from supplier C located in the Waikato region, and nine farm from supplier A located in the Taranaki region (Figure 5.2a). The number of selected samples from each supplier was unevenly distributed with 57.5% (96/167) of the isolates sampled from supplier A, 35.9% (60/167) of isolates sampled from supplier B and 6.6% (11/167) sampled from supplier C.

5.4.2. Whole-genome MLST analysis

A rooted maximum-likelihood phylogenetic tree reconstructed from core polymorphic loci is shown in Figure 5.2b. In this analysis, 230 polymorphic sites were identified including 204 SNP loci shared by the 167 isolates. Within the phylogeny, there were two distinct genetic clusters, with the majority of isolates sampled from farms belonging to the same poultry supplier grouping within the same genetic cluster (Figure 5.3a) and network community (Figure 5.3b) with the exception of a few single isolates that cluster with isolates sampled from farms belonging neither to the same parent company or community. However, the tanglegram generated from the rooted maximum-likelihood phylogenetic tree and hierarchical dendrogram based on Euclidean distance show that these single isolates are located within the same geographical region as other isolates in the cluster (Figure 5.4). Despite this, there was only a weak correlation between the phylogenetic distances and geographical distances with a Spearman ρ value equal to 0.387 for Euclidean distance, and 0.385 for road distance. The tanglegram generated from the rooted maximum-likelihood phylogenetic tree and hierarchical dendrogram based on road distance is shown in Appendix D, Figure D3.

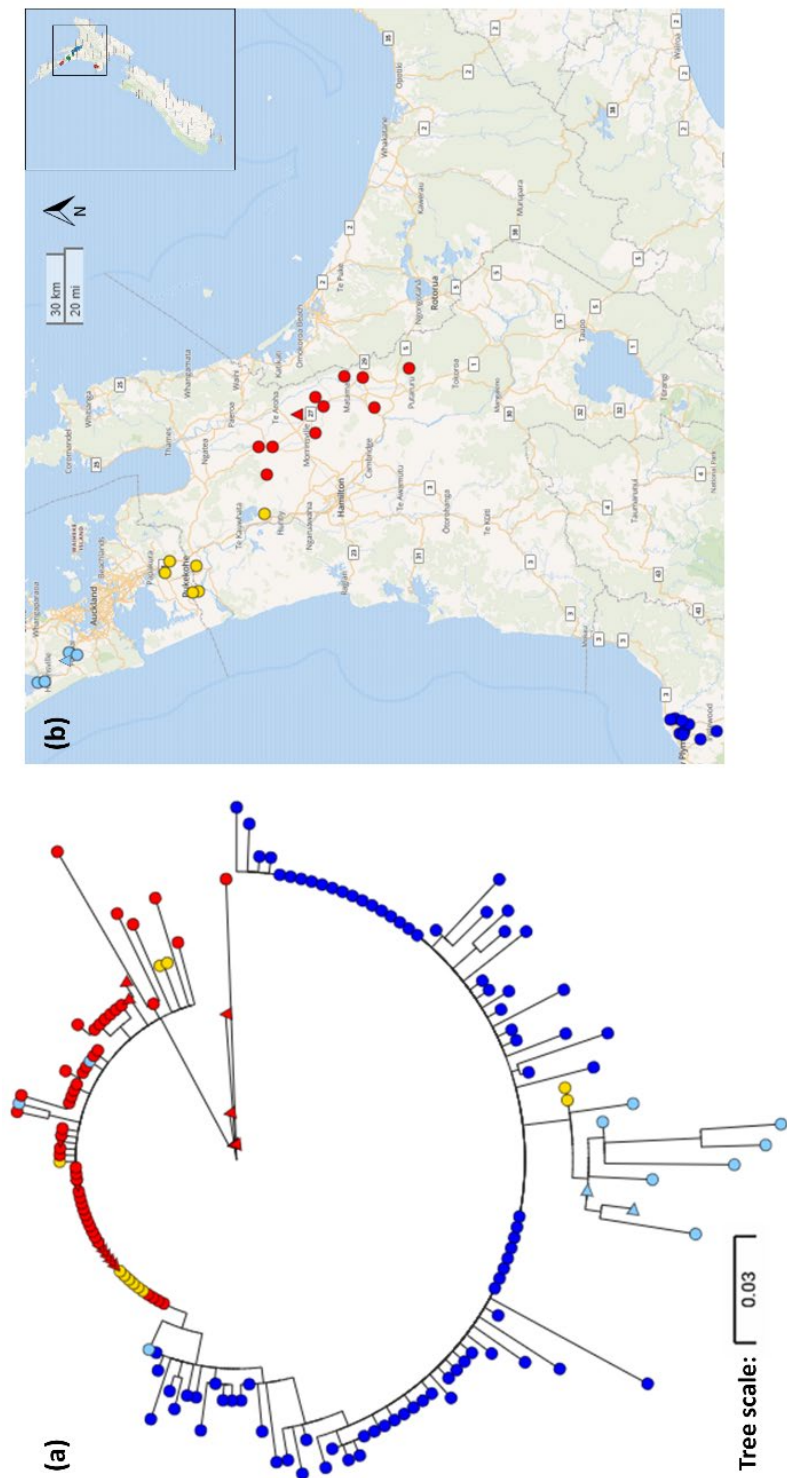


Figure 5.2. (a) Map showing the geographical distribution of New Zealand commercial poultry premises ($n = 30$) with sampled isolates positive for *C. jejuni* ST-6964 ($n = 167$) alongside (b) a phylogeny showing the genetic relatedness of the isolates based on core polymorphic loci. Production type is indicated by the shape of the terminal node (triangles: poultry breeder and circles: poultry grower) with the three colours indicating the three different parent companies to which the farm belongs. The figure has been created using the online tool Microreact (Argimón *et al.* 2016).

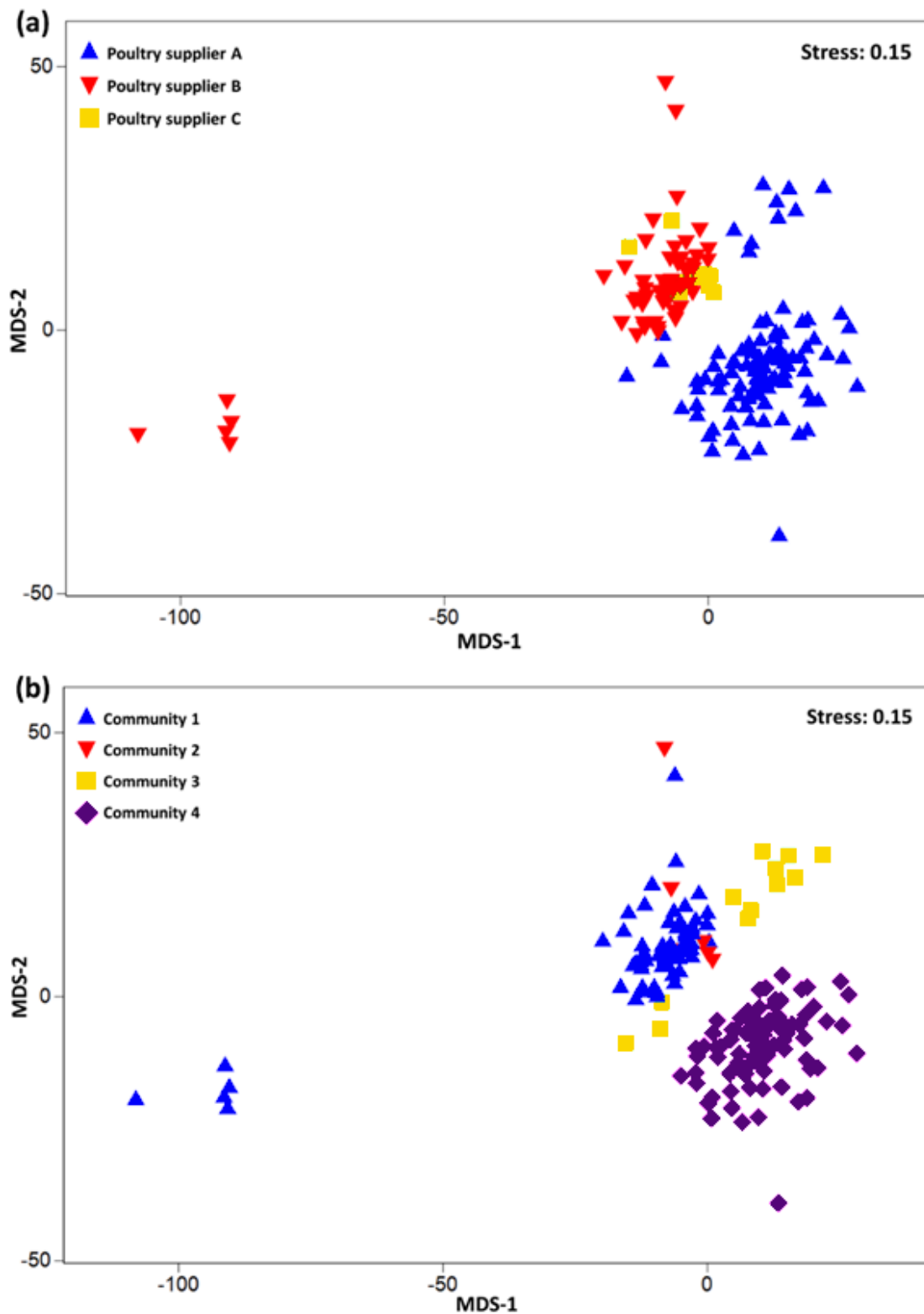


Figure 5.3. Two-dimensional metric multidimensional scaling (*m*MDS) ordination plots based on the allelic dissimilarity matrix between 167 *C. jejuni* isolates with isolates coloured according to (a) the parent company and (b) the network community.

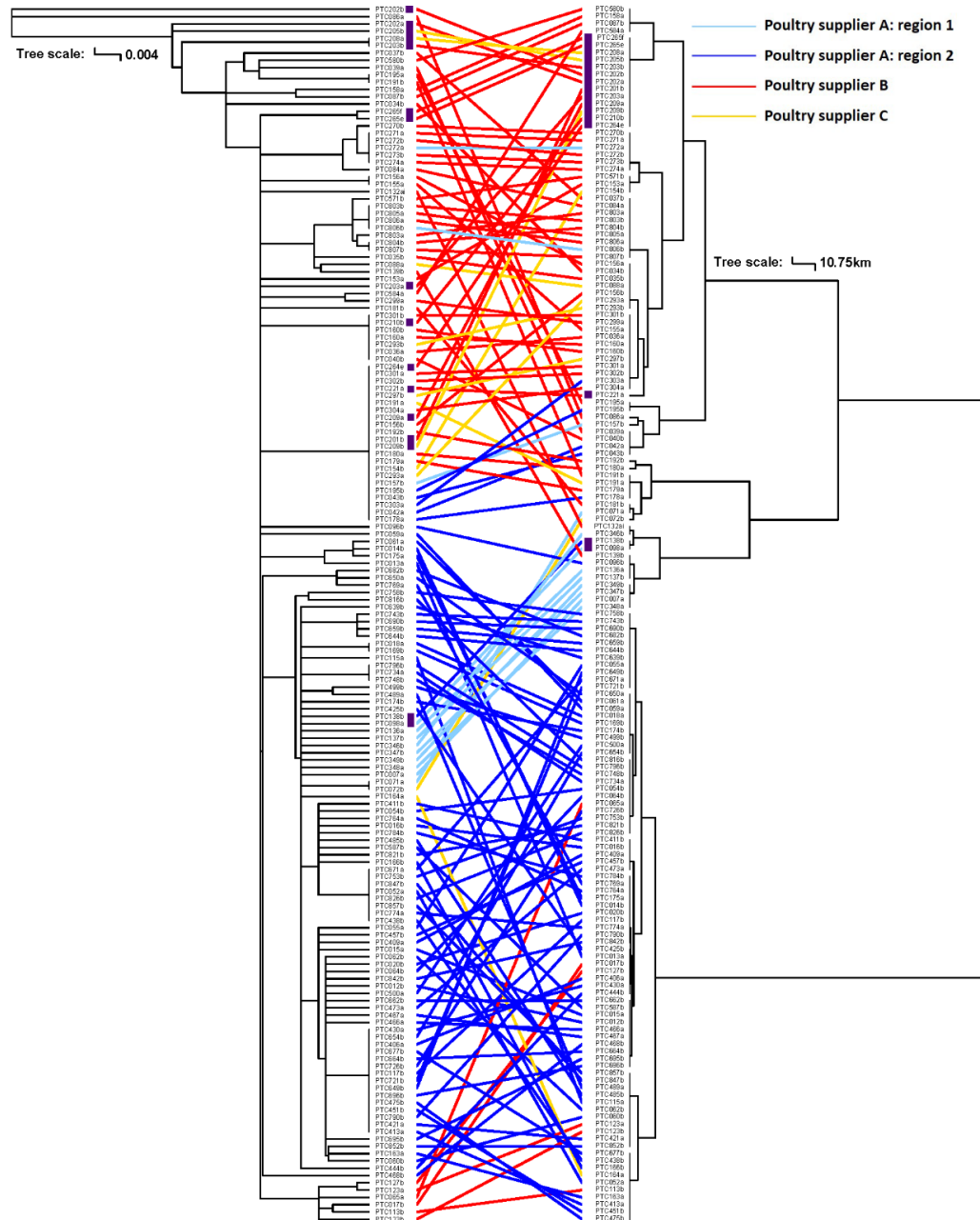


Figure 5.4. A tanglegram containing a maximum-likelihood phylogenetic tree rooted by outgroup showing the population structure of 167 *C. jejuni* ST-6964 isolates based on the core polymorphic loci alignment following core-genome MLST (left) compared to a dendrogram representing the Euclidean distances between the farms from which isolates were sampled (right). The colour of the connecting line indicates the poultry supplier of each farm (A, B, or C) with farms belonging to supplier A located in two geographical regions (regions 1 and 2) in comparison to poultry suppliers B and C whose farms are geographically clustered in one region. A purple square following the isolate ID indicates the farm is a poultry breeder whilst no square indicates it is a poultry grower.

5.4.3. Comparative genomics of mobile elements

Overall, genes from three out of the five *CJIE* reference sequences were distributed differentially among the isolates, with *CJIE1* genes present in 100% of the isolates, *CJIE1v* genes present in 85.0% (142/167) of the isolates and *CJIE4* genes present in 76.6% (128/167) of the isolates (Figure 5.5). The plasmid *pTetO*, was identified in 49.1% (82/167) of the isolates across all poultry suppliers whilst the plasmid *pVir*, was absent from all isolates (Figure 5.6). The functional gene *glcD* was found to be truncated in 31.1% (52/167) of the isolates, all of which were sampled from farms belonging to a single poultry supplier. The *amtB* gene was also truncated in all of these isolates with the addition of three isolates also belonging to the same supplier but with an intact *glcD* gene (Figure 5.6).

5.4.4. Characteristics of model matrices

The allelic distance between isolates ranged from 0-129.78 (mean = 26.55) whilst the *p*-distance ranged from 0-0.16 (mean = 0.04). In the geographical distance matrices, the Euclidean distance between the farms ranged from 1.31km - 282.08km (mean = 131.73km) whilst the road distance ranged from 1.32km - 418.79km (mean = 180.98km). In the network distance matrices based on the imputed networks, no shortest path between any two farms was greater than 3; however, many farms remained unconnected. For example, in the feed network, 59.1% of the potential pathways between farms did not exist, while 39.8% of the pathways were a direct link between two farms. Both the live bird and waste networks had 43.4% of the pathways that were non-existent, while 33.8% and 36.3% were direct links between two farms, respectively. This suggests that in these networks, a greater number of farms were connected, although not so many with direct links. The network graphs, shown in Appendix D, Figure D1, also indicated that live bird and waste networks were more cohesive, having larger betweenness centrality scores (Table 5.2). The high proportion of farms directly linked within the feed network reflects the small number of transporting companies operating within the network. The network

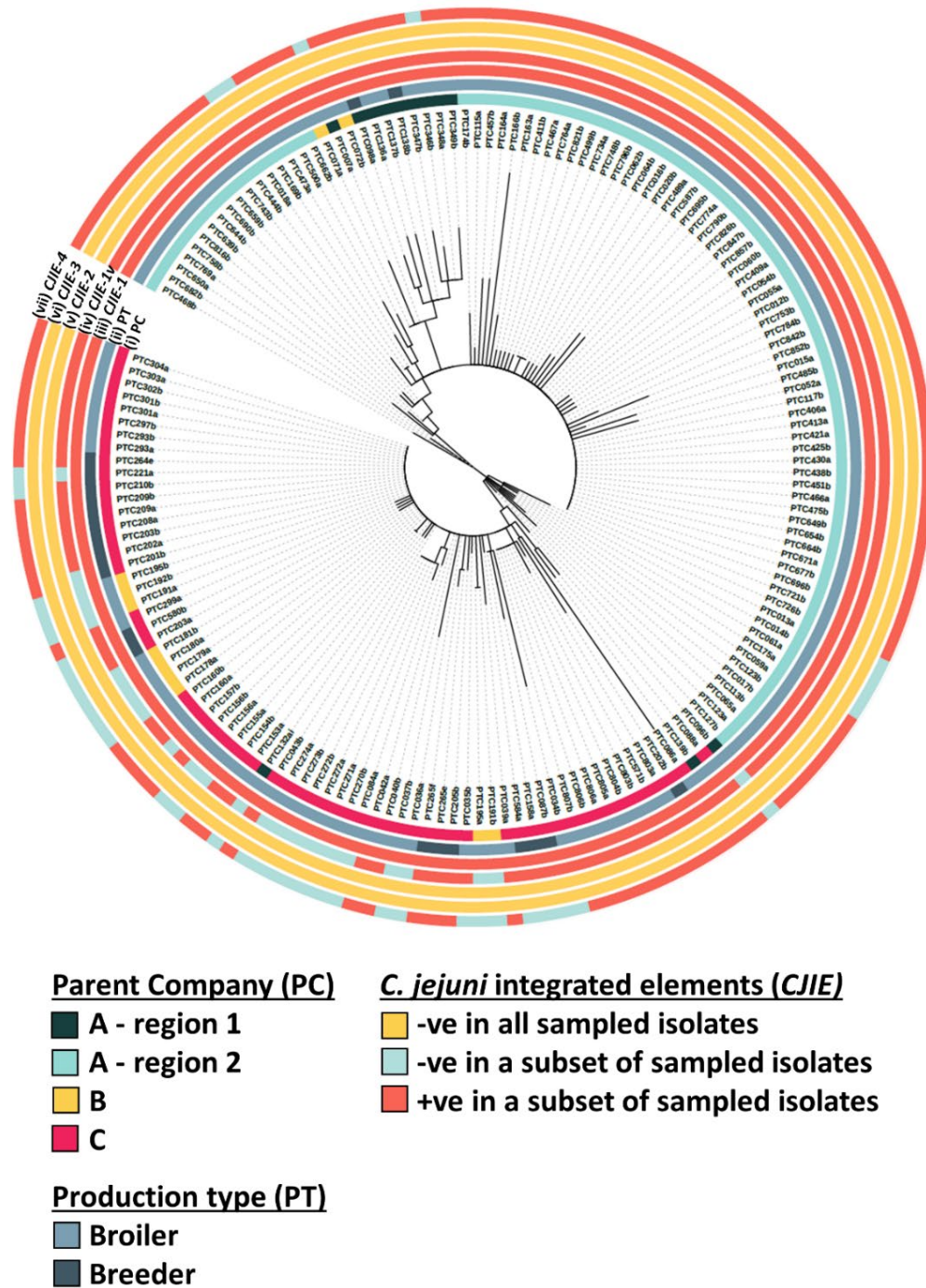


Figure 5.5. A rooted maximum-likelihood phylogenetic tree reconstructed from core polymorphic loci between 167 *C. jejuni* ST-6964 sampled from 30 New Zealand commercial poultry farms. The coloured rings indicate (i) the parent company (PC) of the farm from which the isolate was collected, (ii) the farm production type (PT) (broiler versus breeder), (iii) the presence or absence of *C. jejuni* integrated element 1 (CJIE1), (iv) CJIE1v, (v) CJIE2, (vi) CJIE3, and (vii) CJIE4. The figure has been created using the online tool Interactive Tree of Life (iTOL) (v4.4.2).

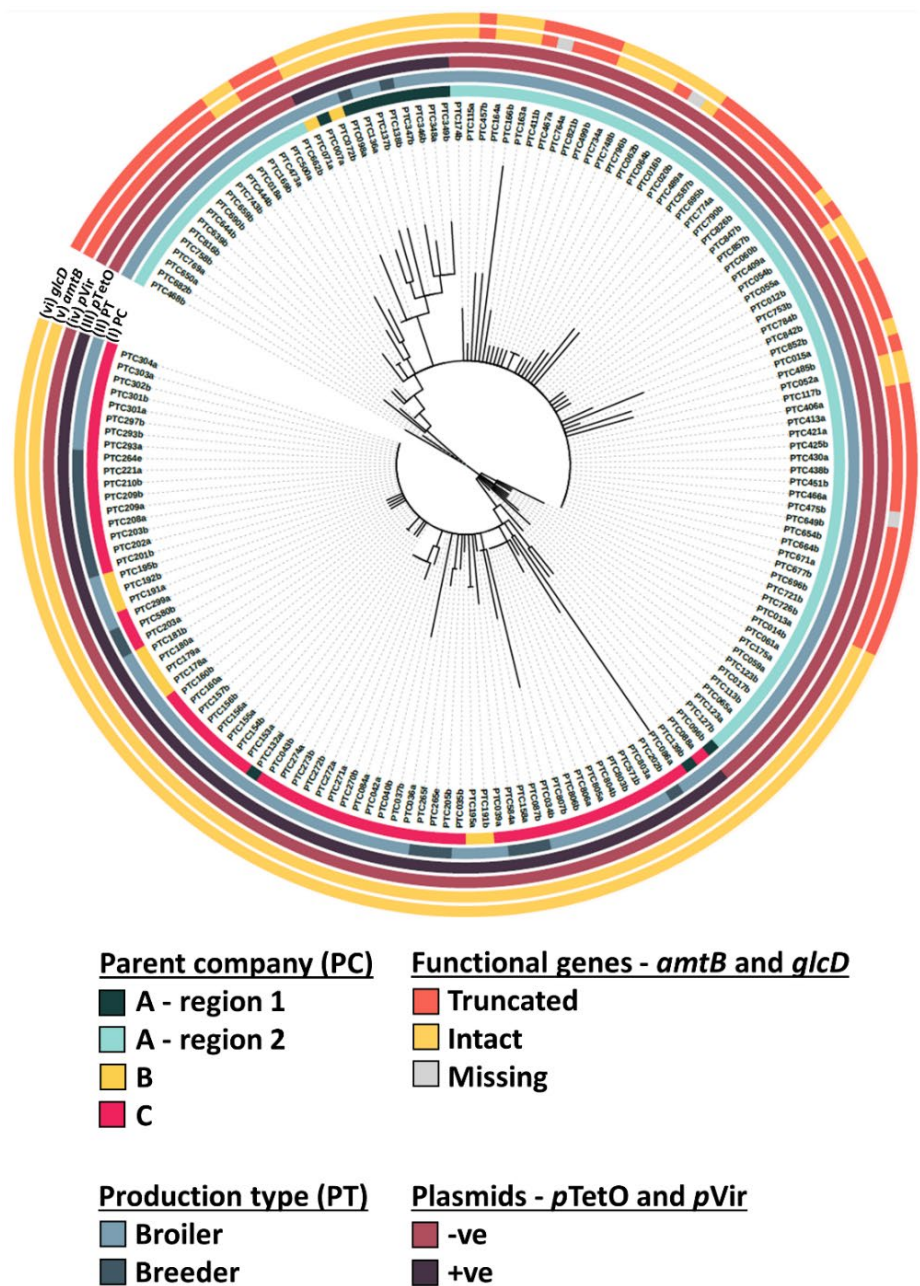


Figure 5.6. A rooted maximum-likelihood phylogenetic tree reconstructed from core polymorphic loci between 167 *C. jejuni* ST-6964 sampled from New Zealand commercial poultry farms (n = 30). The coloured rings indicate (i) the parent company (PC) of the farm from which the isolate was collected (ii) the farm production type (PT) (broiler versus breeder), (iii) the presence or absence of a *Tet*-like plasmid that has been associated with resistance (*pTetO*), (iv) the presence or absence the virulent plasmid *pVir*, and the selected truncation of two functional genes (v) *amtB* encoding an ammonium transporter and (vi) *gltD* encoding the putative glycolate oxidase subunit D. The figure has been created using the online tool Interactive Tree of Life (iTOL) (v4.4.2).

Table 5.2. Network statistics describing both the empirical network (*i.e.*, the network constructed from 16 survey responses reporting all on- and off-farm movements over a one-year period) and the imputed network (*i.e.*, the network constructed using all 30 farms from which isolates were sampled, with missing edges imputed based on expert opinion) constructed from all on- and off-farm movements relating to (i) feed, (ii) live birds and hatching eggs or (iii) waste and litter within the New Zealand commercial poultry industry. Network metrics include the “Degree”, indicating the total number of on- and off-farm movements on a single farm in the network and, “Betweenness” indicating the frequency a farm is on the shortest path between any two other farms in the network.

Network metric	Network	Empirical network	Imputed network
Number of nodes	<i>All</i>	16	30
Number of Edges	<i>Feed</i>	208	455
	<i>Live birds & hatching eggs</i>	457	1121
	<i>Waste & litter</i>	367	776
	<i>Combined</i>	1032	2352
Mean degree^a (min-max)	<i>Feed</i>	6.25 (0-8)	11.53 (3-16)
	<i>Live birds & hatching eggs</i>	57.12 (4-94)	74.73 (18-109)
	<i>Waste & litter</i>	5.63 (1-8)	10.53 (4-17)
	<i>Combined</i>	6.75 (3-8)	13.73 (8-20)
Mean betweenness^b (min-max)	<i>Feed</i>	0.06 (0-0.25)	0.17 (0-0.45)
	<i>Live birds & hatching eggs</i>	0.56 (0-4)	3.57 (0-90)
	<i>Waste & litter</i>	1.13 (0-9)	3.13 (0-28)
	<i>Combined</i>	0.19 (0-1)	1.33 (0-6)

graphs showed a clustering of farms around each of the 3 companies operating within the feed network with no links between the clusters (Appendix D, Figure D1). In contrast, the live bird and waste networks had a greater number of links between clusters (*i.e.*, individual farms are using more than one transporting company). The community analysis performed on the imputed network combining all the on- and off-farm movements identified four communities (Appendix D, Figure D2) with 56.7% (17/30) of the farms belonging to the largest community. The *m*MDS plot, based on the allelic

genetic distance matrix suggested that pathogens in the same network community have similar genetic structures (Figure 5.3b).

5.4.5. Relating genetic distances and model matrices

All model matrices showed a significant relationship ($p < 0.001$) with the unconstrained allelic dissimilarity matrix. The Spearman rank matrix correlation coefficients ranged from 0.438 to 0.632 (Table 5.3). In addition, parent company had a significant effect on the genetic structure of isolates, and there was also significant variation due to individual farms (Appendix D, Table D1). After removing the effects of farm and parent company, only the feed network had a significant matrix correlation with the residual allelic dissimilarity matrix ($p < 0.05$). The second stage *m*MDS plot showed that Euclidean distance and road distance matrices were very highly correlated, as were the live bird and waste networks, whilst minimum spanning trees (MSTs) indicated that the feed network had the closest relationship with the allelic dissimilarities among pathogen isolates (Figure 5.7).

Table 5.3. Spearman's rank matrix correlation (ρ) between each model matrix and (i) the allelic dissimilarity matrix between 167 *C. jejuni* isolates and (ii) the residual allelic dissimilarity matrix after fitting the ANOVA factors of parent company ($n = 3$) and farm ($n = 30$) nested within parent company, with p -values obtained using 9999 permutations.

	(i) Unconstrained matrix		(ii) Residual matrix	
	ρ	p -value	ρ	p -value
Feed	0.632	0.0001	0.040	0.0370
Live birds	0.623	0.0001	0.001	0.4154
Waste	0.614	0.0001	0.005	0.3336
Road distance	0.584	0.0001	-0.005	0.5491
Euclidean distance	0.580	0.0001	-0.024	0.8023
Parent company	0.438	0.0001	-	-

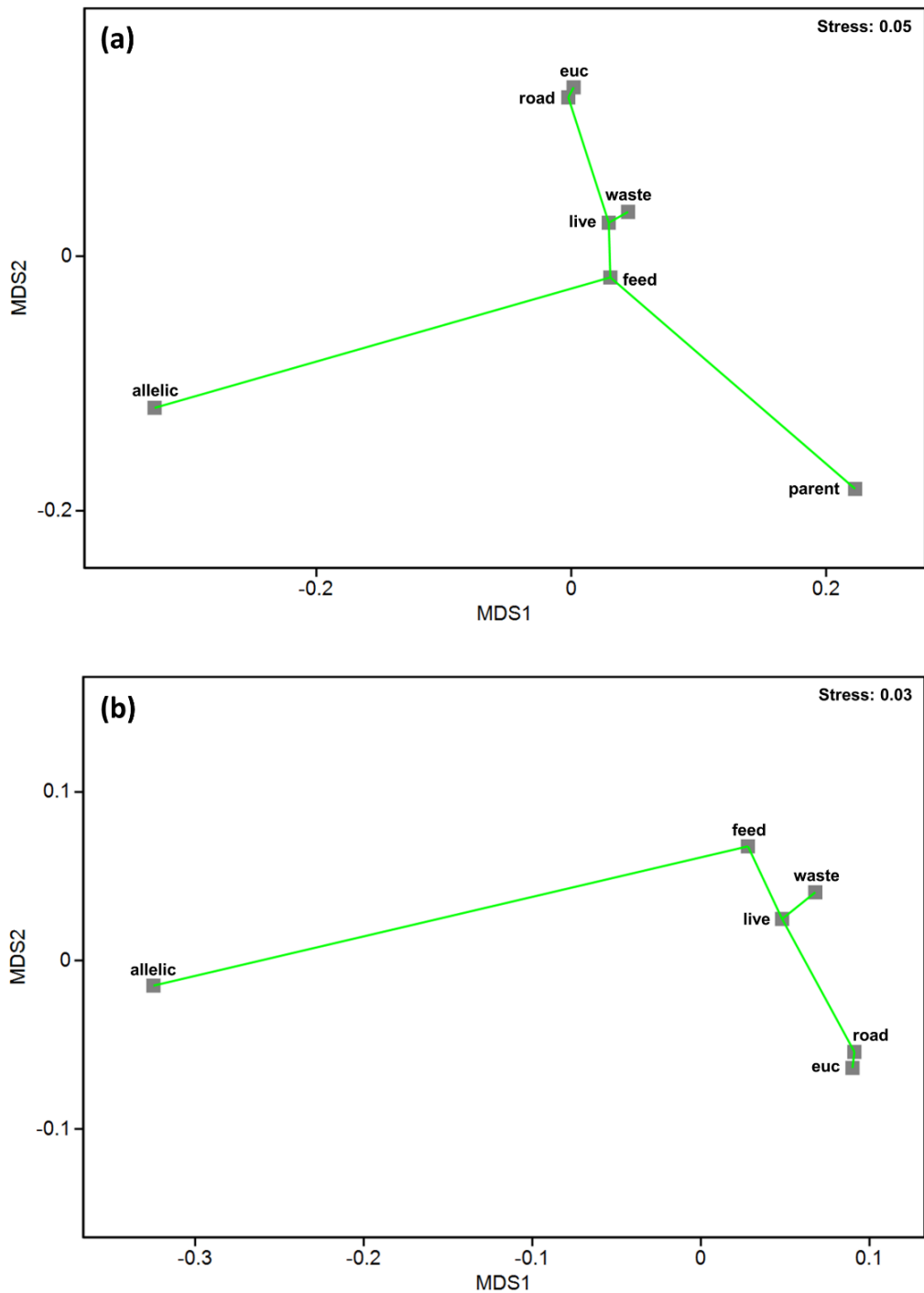


Figure 5.7. Second-stage metric multidimensional scaling (*mMDS*) ordination plots showing proximities (rank matrix correlations) between the allelic dissimilarity matrix (“allelic”) and each of several model distance matrices: “euc” = Euclidean distance; “road” = road distance; “feed” = feed network distance; “live” = live birds network distance; “waste” = waste and litter network distance; “parent” = parent company distance. Superimposed on each plot is the minimum spanning tree (MST). The *mMDS* plot shown in (a) includes parent company whereas (b) does not include parent company.

5.4.6. Explaining variation in genetic distances

When considered individually (marginal tests), road distance accounted for the largest proportion of the variation in the allelic dissimilarity matrix, followed by the Euclidean distance and the feed network (Table 5.4). Sequential tests showed that, after the addition of road distance, farm, parent company, and feed network, the remaining networks added little to explain genetic variation (Table 5.5; $p > 0.50$). Indeed, the model with the lowest AIC value (out of the class of models that included per force the parent company and the farm) included only feed network and road distance (Table 5.6). A similar result was found for the empirical networks (Appendix D, Tables D4-D8). However, there was a slight discrepancy in the models built using the alternative genetic distance measure, the p -distance, which did not include the feed network in the model with the best AIC value (Appendix D, Tables D9-D18).

Table 5.4. Individual distance-based redundancy analysis (dbRDA) models to explain variation in allelic dissimilarities among 167 *C. jejuni* isolates in response to each of two ANOVA factors (parent company or farm nested in parent company) or sets of regression coordinates corresponding to geographic position (Euclidean distance or road distance) or the network model matrices of interest (feed, live birds or waste), with p -values for each of these separate marginal tests obtained using 9999 unrestricted permutations. “Prop” gives the proportion of the total variation explained whilst “ df ” gives the numerator (regression) and denominator (residual) degrees of freedom for the test. Models have been presented in order of decreasing R^2 values.

	Prop.	Pseudo- F	df	p -value
Parent Company	0.2649	29.55	3, 164	0.0001
Waste	0.2854	32.76	3, 164	0.0001
Farm	0.2894	2.10	28, 139	0.0117
Live birds	0.2912	33.69	3, 164	0.0001
Feed	0.3184	25.39	4, 163	0.0001
Euclidean distance	0.3458	43.33	3, 164	0.0001
Road distance	0.4112	37.95	4, 163	0.0001

Table 5.5. Distance-based redundancy analysis (dbRDA) to explain variation in allelic dissimilarities among 167 *C. jejuni* isolates in response to the factors and sets of regressors listed in Table 5.4, but here conditional tests were done in a sequential stepwise fashion under forward selection based on R^2 . Each test used 9999 permutations of residuals under a reduced model. “Prop” gives the proportion of additional variation explained by adding that set of variables to the model, “Cumul” tracks the cumulative explained variation with each added step, and “ df ” provides the regression and residual degrees of freedom. Note: from step 5 onwards, additional explained variation is < 1%.

Step		Prop.	Cumul.	df	Pseudo- F	p -value
1	+Road	0.4112	0.4112	4, 163	37.95	0.0001
2	+Farm	0.1475	0.5588	31, 136	1.684	0.0438
3	+Parent company	0.0156	0.5743	33, 134	2.453	0.0497
4	+Feed	0.0192	0.5935	36, 131	2.063	0.0839
5	+Live birds	0.0052	0.5987	38, 129	0.8342	0.5017
6	+Euclidean	0.0030	0.6018	40, 127	0.4805	0.6538
7	+Waste	0.0025	0.6042	42, 125	0.3891	0.8696

Table 5.6. Top four dbRDA models obtained on the basis of the multivariate analogue to the Akaike Information Criterion (AIC) (see text) to explain variation in allelic dissimilarities among 167 *C. jejuni* isolates. Two ANOVA factors (parent company and farms nested in parent company) were included in all potential models *a priori*.

Model selections	R^2	No. Sets	AIC
Feed network, road distance, parent company and farm	0.5935	4	975.23
Road distance, parent company and farm	0.5743	3	976.94
Feed network, live bird network, road distance, parent company and farm	0.5987	5	977.09
Euclidean distance, feed network, road distance, parent company and farm	0.5964	5	978.06

5.5. Discussion

This study was one of the first to combine network data with molecular data in PERMANOVA and DistLM analyses to identify transmission pathways important to the spread of *C. jejuni* in the New Zealand poultry industry. Overall, study results support previous findings showing an association between poultry suppliers and pathogen phylogeny (McTavish *et al.* 2008; Müllner *et al.* 2009, 2010) with the genetic distance between isolates from different poultry suppliers greater than the distance between isolates from the same poultry supplier. However, this study also found a significant association between the road distance, Euclidean distance and the network distance of transporting feed vehicles with the genetic relatedness of the 167 *C. jejuni* ST-6964 isolates from across the 30 poultry farms. These results suggest that the transportation of feed within the commercial poultry industry may play a significant role in the spread of *C. jejuni* between flocks however, it is not clear if it is the transporting vehicle that is important or the feed itself that may be acting as a vector. Similarly, results suggest that local contacts may also play a role however, without further information the exact mechanism cannot be deduced with many local pathways known to contribute to the spread of *C. jejuni*; such as the movement of wildlife or personnel.

This study also determined the presence of three *CJIEs*, genomic integrated elements identified in *C. jejuni* strain RM1221 (Fouts *et al.* 2005), in the New Zealand isolates. This demonstrates a high genetic diversity between the study *C. jejuni* isolates, all of which belong to ST-6964, and highlights how the presence of bacteriophage-related genes in integrated elements are contributing to the genetic diversity seen in *C. jejuni*. Furthermore, the similarity between the presence-absence profile of both *CJIE1v* and *CJIE4* suggests that these integrated elements may not only help to differentiate between closely related strains but also help determine epidemiological links between isolates belonging to the same strain. The presence-absence of the *pTet* plasmid also contributes to the genetic differentiation between the study isolates; although, it was identified across

all three poultry suppliers in multiple regions making it difficult to determine transmission dynamics, especially if there is limited evidence that isolates are epidemiologically related. Lastly, this study examined the state of two functional genes, *amtB* and *glcD*, that have previously been associated with human infection when intact. Both genes were found to be truncated in isolates from a single poultry supplier and only within one region. Although it is not possible to determine the clinical relevance of this selective truncation, the relatively limited number of observed truncated genes suggests that having an intact *amtB* and *glcD* gene may offer a selective advantage.

Although our study has potential implications for future surveillance and control activities, it did have limitations, including a relatively small number of study farms and sampled isolates. Only 39.2% (62/158) of commercial broiler producers registered with PIANZ or EPF were sampled for this study. Having missing farms made it difficult to represent the true distribution of *C. jejuni* ST-6964 across New Zealand. Furthermore, only a limited number of samples was taken from each farm, increasing the chances of there being farms where *C. jejuni* ST-6964 was present but not detected. A small sample size is also likely to hinder phylogenetic reconstruction (Nabhan and Sarkar, 2012). Although this is more of a problem when attempting to infer transmission between farms, a greater resolution could have been gained by sampling a greater number of farms or by sequencing more *C. jejuni* positive isolates from the sampled farms in this study. Lastly, by only constructing networks using the small proportion of broiler farms that were sampled here, artefacts may arise in community analysis: farms identified within a community may only belong to that community when considering a particular subset of network nodes and edges (Shizuka and Farine, 2016).

Another limitation of our study was the low response rate to the industry survey, hence the potential for both non-response bias and recall bias to occur. Only 53.3% (16/30) of the farms in the study had movement data that could be used to construct the contact

networks, and although movement data for non-responders was collected based on expert opinion, it is hard to quantify the impact of this non-response bias. However, the similarity in results obtained using the imputed and the empirical networks provided some reassurance regarding the network imputation methods used. Secondly, by using a survey questionnaire the study results are likely to suffer from limitations inherent to most questionnaires, such as recall bias, with many previous studies showing a poor correlation between survey responses and on-farm practices (Sax *et al.* 2003; Bewsell, 2010; Racicot *et al.* 2012). In this study, producers were asked to recall all on- and off-farm movements over a one-year period. This was thought to be appropriate due to the steady nature of the poultry industry and the relationship between parent company and on-farm contractors, making it easier for producers to name all the transporting companies that deliver goods or services such as feed. However, over a one-year period, it may be hard for producers to recall more informal movements, such as the on- and off-farm movement of farm visitors, or irregular movements such as the redistribution of left-over feed at the end of a production cycle.

In addition to the effects of network imputation, this study also looked at the effect of using two different genetic distance measures on the relationship between the model matrices and the genetic relatedness of the *C. jejuni* isolates. The discrepancies between the final models fitted to different genetic distance dissimilarity matrices creates some doubt as to the exact contribution of transporting feed vehicles in the spread *C. jejuni* ST-6964. These discrepancies may be explained, however, by the resolution captured within each of the matrices. Overall, there should be greater resolution within the allelic dissimilarity matrix that was produced using a whole-genome multi-locus sequence typing (wg-MLST) and therefore includes homopolymeric tracts; repetitive DNA sequences that can be used to distinguish between very closely related isolates, such as those within the same ST as in this study (Zhang *et al.* 2015; Allard *et al.* 2016; Schürch *et al.* 2018). In contrast, the *p*-distance matrix was produced using a SNP-based approach,

looking at just the core genome in which recombinant regions had been removed. This means that only core SNPs were used to distinguish between isolates, meaning that some resolution may have been lost, especially considering the small timeframe over which isolates were collected.

It is clear that more research is needed to understand why and how these contact networks may be important for the spread of *C. jejuni*. In particular, being able to identify the attributes within the feed network and the pathways captured by both road distance and Euclidean distance contributing to transmission would help inform targeted control activities within the commercial poultry industry. Future research will rely on both the participation of producers and innovative technology to capture and track all on- and off-farms movements that may be contributing to the spread of disease. For example, Global Positioning Systems (GPS) may be used to track the spatial pattern of transporting vehicles within livestock contact networks, or new recording systems may be developed to aid the producer to track all vehicles entering their farm, such as the OnSide[®] smartphone-based app (<http://www.onside.co.nz/>). Further research into on-farm biosecurity is also important to see if producers are identifying these pathways as major risks for disease transmission and if they are implementing any disease risk management strategies to target these pathways such as the disinfection of vehicles. Without this additional research, it will be difficult to recommend suitable control strategies that may target pathways contributing to the spread of *C. jejuni* within the commercial poultry industry.

5.6. Conclusion

This study highlights the importance of both local transmission mechanisms and the movement of transporting vehicles in the spread of *C. jejuni* within the commercial poultry industry. This includes not only transmission between farms sharing the same parent company but also between farms in different supply chains. Although this spread

is likely to be facilitated by the small number of companies servicing the industry, particularly those few delivering and transporting feed, additional research is required to fully characterise these risk pathways and to develop appropriate control strategies that would reduce the spread of *C. jejuni* within the commercial poultry industry.

5.7. Acknowledgements

The authors are grateful for the support provided by the staff of the Poultry Industry Association and the Egg Producers' Federation of New Zealand. We would also like to thank all the poultry producers who took the time to complete the survey questionnaire. This study was supported by the Ministry for Primary Industries, Massey University School of Veterinary Science, and the Biological Heritage National Science Challenge. In addition, JZ and NPF was funded by the New Zealand Food Safety Science & Research Centre.

5.8. References

- Ahn Y-Y, Bagrow JP, Lehmann S. Link communities reveal multiscale complexity in networks. *Nature* 466, 761–764, 2010
- Allard MW, Strain E, Melka D, Bunning K, Musser SM, Brown EW, Timme R. Practical value of food pathogen traceability through building a whole-genome sequencing network and database. *J. Clin. Microbiol.* 54, 1975–1983, 2016
- Anderson MJ, Gorley RN, Clarke KR. *PERMANOVA+ for PRIMER: guide to software and statistical methods*. PRIMER-E, Plymouth, UK, 2008
- Anderson MJ. *Permutational Multivariate Analysis of Variance (PERMANOVA)*. In: Wiley StatsRef: Statistics Reference Online. Pg. 1–15. John Wiley & Sons, Ltd., Chichester, UK, 2017
- Antonovics J. Transmission dynamics: critical questions and challenges. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372, e20160087, 2017
- Antonovics J, Wilson AJ, Forbes MR, Hauffe HC, Kallio ER, Leggett HC, Longdon B, Okamura B, Sait SM, Webster JP. The evolution of transmission mode. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372, e20160083, 2017
- Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, Feil EJ, Holden MTG, Yeats CA, Grundmann H, Spratt BG, Aanensen DM. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb. Genom.* 2, e000093, 2016
- Bacon DJ, Alm RA, Hu L, Hickey TE, Ewing CP, Batchelor RA, Trust TJ, Guerry P. DNA sequence and mutational analyses of the *pVir* plasmid of *Campylobacter jejuni* 81-176. *Infect. Immun.* 70, 6242–6250, 2002
- Baines SL, Howden BP, Heffernan H, Stinear TP, Carter GP, Seemann T, Kwong JC, Ritchie SR, Williamson DA. Rapid emergence and evolution of *Staphylococcus aureus* clones harboring fusc-containing staphylococcal cassette chromosome elements. *Antimicrob. Agents Chemother.* 60, 2359–2365, 2016
- Bajardi Paolo, Barrat Alain, Savini Lara, Colizza Vittoria. Optimizing surveillance for livestock disease spreading through animal movements. *J. R. Soc. Interface* 9, 2814–2825, 2012

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prijbelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477, 2012
- Bewsell D. *Determining individuals' response to New Zealand biosecurity*. Degree of Master of Commerce and Management. Lincoln University, Lincoln, New Zealand, 2010
- Boto L. Horizontal gene transfer in evolution: facts and challenges. *Proc. Biol. Sci.* 277, 819–827, 2010
- Chaters GL, Johnson PCD, Cleaveland S, Crispell J, de Glanville WA, Doherty T, Matthews L, Mohr S, Nyasebwa OM, Rossi G, Salvador LCM, Swai E, Kao RR. Analysing livestock network data for infectious disease control: an argument for routine data collection in emerging economies. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* 374, doi: 10.1098/rstb.2018.0264, 2019
- Christley RM, Pinchbeck GL, Bowers RG, Clancy D, French NP, Bennett R, Turner J. Infection in social networks: using network analysis to identify high-risk individuals. *Am. J. Epidemiol.* 162, 1024–1031, 2005
- Clarke K, Gorley R. *PRIMER v7: User Manual/Tutorial*. PRIMER-E, Plymouth, UK, 2015
- Craft ME. Infectious disease transmission and contact networks in wildlife and livestock. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, doi: 10.1098/rstb.2014.0107, 2015
- Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* 43, e15, 2015
- Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, Complex Systems*, 2006. Available at: <http://igraph.org>
- Duong T, Konkel ME. Comparative studies of *Campylobacter jejuni* genomic diversity reveal the importance of core and dispensable genes in the biology of this enigmatic food-borne pathogen. *Curr. Opin. Biotechnol.* 20, 158–165, 2009
- Enns EA, Brandeau ML. Inferring model parameters in network-based disease simulation. *Health Care Manag. Sci.* 14, 174–188, 2011

- Eubank S, Guclu H, Anil Kumar VS, Marathe MV, Srinivasan A, Toroczkai Z, Wang N. Modelling disease outbreaks in realistic urban social networks. *Nature* 429, 180–184, 2004
- Fouts DE, Mongodin EF, Mandrell RE, Miller WG, Rasko DA, Ravel J, Brinkac LM, DeBoy RT, Parker CT, Daugherty SC, Dodson RJ, Durkin AS, Madupu R, Sullivan SA, Shetty JU, Ayodeji MA, Shvartsbeyn A, Schatz MC, Badger JH, Fraser CM, Nelson KE. Major structural differences and novel potential virulence mechanisms from the genomes of multiple *Campylobacter* species. *PLoS Biol.* 3, e15, 2005
- Freedman D, Lane D. A non-stochastic interpretation of reported significance levels. *J. Bus. Econ. Stat.* 1, 292–298, 1983
- French NP, Zhang J, Carter GP, Midwinter AC, Biggs PJ, Dyet K, Gilpin BJ, Ingle DJ, Mulqueen K, Rogers LE, Wilkinson DA, Greening SS, Mullner P, Williamson DA. Genomic analysis of fluoroquinolone- and tetracycline-resistant *Campylobacter jejuni* sequence type 6964 in humans and poultry, New Zealand, 2014–2016 *Emerg. Infect. Dis.* 25, 2226–2234, 2019
- Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. *Software-Practice and Experience* 21, 1129–1164, 1991
- Galili T. dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 31, 3718–3720, 2015
- Gates MC, Woolhouse MEJ. Controlling infectious disease through the targeted manipulation of contact network structure. *Epidemics* 12, 11–19, 2015
- Gilbertson MLJ, Fountain-Jones NM, Craft ME. Incorporating genomic methods into contact networks to reveal new insights into animal behaviour and infectious disease dynamics. *Behaviour* 155, 759–791, 2018
- Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, Holmes EC. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303, 327–332, 2004
- Hijmans RJ. geosphere: Spherical Trigonometry. *R package version 1.5-10*. 2019. Available at: <https://CRAN.R-project.org/package=geosphere>

- Jombart T, Cori A, Didelot X, Cauchemez S, Fraser C, Ferguson N. Bayesian reconstruction of disease outbreaks by combining epidemiologic and genomic data. *PLoS Comput. Biol.* 10, e1003457, 2014
- Kaakoush NO, Castaño-Rodríguez N, Mitchell HM, Man SM. Global epidemiology of *Campylobacter* infection. *Clin. Microbiol. Rev.* 28, 687–720, 2015
- Kalinka AT, Tomancak P. linkcomm: An R package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type. *Bioinformatics* 27, 2011–2012, 2011
- Kingman JFC. Origins of the Coalescent: 1974–1982. *Genetics* 156, 1461–1463, 2000
- Koutsoumanis KP, Lianou A, Gougouli M. Latest developments in foodborne pathogens modeling. *Curr. Opin. Food Sci.* 8, 89–98, 2016
- Kruskal JB, Wish M. *Multidimensional Scaling*. Sage Publications, Newbury Park, California, US, 1978.
- Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874, 2016
- Letunic I, Bork P. Interactive Tree of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23, 127–128, 2007
- Legendre P, Anderson MJ. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol. Monogr.* 69, 1–24, 1999
- Leventhal GE, Kouyos R, Stadler T, von Wyl V, Yerly S, Böni J, Cellerai C, Klimkait T, Günthard HF, Bonhoeffer S. Inferring epidemic contact structure from phylogenetic trees. *PLoS Comput. Biol.* 8, e1002413, 2012
- Lin J. Novel Approaches for *Campylobacter* control in poultry. *Foodborne Patho. Dis.* 6, 755–765, 2009
- Llarena A-K, Zhang J, Vehkala M, Välimäki N, Hakkinen M, Hänninen M-L, Roasto M, Mäesaar M, Taboada E, Barker D, Garofolo G, Cammà C, Di Giannatale E, Corander J, Rossi M. Monomorphic genotypes within a generalist lineage of *Campylobacter jejuni* show signs of global dispersion. *Microb. Genom.* 2, 2016

- Lockhart CY, Stevenson MA, Rawdon TG, Gerber N, French NP. Patterns of contact within the New Zealand poultry industry. *Prev. Vet. Med.* 95, 258–266, 2010
- Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27, 209–220, 1967
- McArdle BH, Anderson MJ. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* 82, 290–297, 2001
- McTavish SM, Pope CE, Nicol C, Sexton K, French N, Carter PE. Wide geographical distribution of internationally rare *Campylobacter* clones within New Zealand. *Epidemiol. Infect.* 136, 1244–1252, 2008
- Melo RA, Rodriguez T, Zarruk D. gmapsdistance: Distance and travel time between two points from Google maps. *R package version 3.4.*, 2018. Available at: <https://CRAN.R-project.org/package=gmapsdistance>
- Muellner P, Kells N, Campbell D. *Risk profile the emergence of Campylobacter jejuni st 6964 in poultry in New Zealand and its associated antimicrobial resistance*. MPI Technical Paper No: 2016/16, Wellington, New Zealand, 2016
- Müllner P, Collins-Emerson JM, Midwinter AC, Carter P, Spencer SEF, van der Logt P, Hathaway S, French NP. Molecular epidemiology of *Campylobacter jejuni* in a geographically isolated country with a uniquely structured poultry industry. *Appl. Environ. Microbiol.* 76, 2145–2154, 2010
- Müllner P, Spencer SEF, Wilson DJ, Jones G, Noble AD, Midwinter AC, Collins-Emerson JM, Carter P, Hathaway S, French NP. Assigning the source of human campylobacteriosis in New Zealand: A comparative genetic and epidemiological approach. *Infect. Genet. Evol.* 9, 1311–1319, 2009
- Nabhan AR, Sarkar IN. The impact of taxon sampling on phylogenetic inference: a review of two decades of controversy. *Brief Bioinform.* 13, 122–134, 2012
- Newell DG, Elvers KT, Dopfer D, Hansson I, Jones P, James S, Gittins J, Stern NJ, Davies R, Connerton I, Pearson D, Salvat G, Allen VM. Biosecurity-based interventions and strategies to reduce *Campylobacter* spp. on poultry farms. *Appl. Environ. Microbiol.* 77, 8605–8614, 2011
- New Zealand Food Safety Authority (NZFSA). *Campylobacter risk management strategy*. New Zealand Food Safety Authority, Wellington, New Zealand, 2008

- Olson CK, Ethelberg S, Pelt W van, Tauxe RV.** *Epidemiology of Campylobacter jejuni infections in industrialized nations*. In: Bacterial Pathogenesis in *Campylobacter*, ASM Press, Washington, DC, 2008
- Paradis E, Schliep K.** ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528, 2019
- Racicot M, Venne D, Durivage A, Vaillancourt J-P.** Evaluation of strategies to enhance biosecurity compliance on poultry farms in Québec: Effect of audits and cameras. *Prev. Vet. Med.* 103, 208–218, 2012
- Sah P, Otterstatter M, Leu ST, Leviyang S, Bansal S.** Revealing mechanisms of infectious disease transmission through empirical contact networks. *bioRxiv*, doi: 10.1101/169573, 2018.
- Sax LJ, Gilmartin SK, Bryant AN.** Assessing response rates and nonresponse bias in web and paper surveys. *Res. Higher Educat.* 44, 409–432, 2003
- Schliep KP.** phangorn: Phylogenetic analysis in R. *Bioinformatics* 27, 592–593, 2011
- Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV.** Whole-genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clin. Microbiol. Infect.* 24, 350–354, 2018
- Sears A, Baker MG, Wilson N, Marshall J, Muellner P, Campbell DM, Lake RJ, French NP.** Marked *Campylobacteriosis* decline after interventions aimed at poultry, New Zealand. *Emerg. Infect. Dis.* 17, 1007–1015, 2011
- Seemann T.** Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069, 2014
- Shen Z, Patil RD, Sahin O, Wu Z, Pu X-Y, Dai L, Plummer PJ, Yaeger MJ, Zhang Q.** Identification and functional analysis of two toxin–antitoxin systems in *Campylobacter jejuni*. *Mol. Microbiol.* 101, 909–923, 2016
- Shizuka D, Farine DR.** Measuring the robustness of network community structure using assortativity. *Animal Behaviour* 112, 237–246, 2016

- Silk MJ, Drewe JA, Delahay RJ, Weber N, Steward LC, Wilson-Aggarwal J, Boots M, Hodgson DJ, Croft DP, McDonald RA.** Quantifying direct and indirect contacts for the potential transmission of infection between species using a multilayer contact network. *Behaviour* 155, 731–757, 2018
- Sokal RR, Rohlf FJ.** The comparison of dendrograms by objective methods. *Inter. Assoc. Plant Taxon.* 11, 33–40, 1962
- Somerfield P, Clarke K.** Taxonomic levels, in marine community studies, revisited. *Mar. Ecol. Prog. Ser.* 127, 113–119, 1995
- Stadler T, Bonhoeffer S.** Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368, e20120198, 2013
- Stafford K.** *Livestock Production in New Zealand.* Massey University Press, Auckland, New Zealand, 2017
- Tracz DM, Keelan M, Ahmed-Bentley J, Gibreel A, Kowalewska-Grochowska K, Taylor DE.** *pVir* and bloody diarrhoea in *Campylobacter jejuni*. *Emerg. Infect. Dis.* 11, 839–843, 2005
- Webster Joanne P., Borlase Anna, Rudge James W.** Who acquires infection from whom and how? Disentangling multi-host and multi-mode transmission dynamics in the ‘elimination’ era. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372, e20160091, 2017
- Williamson DA, Dyet K, Heffernan H.** *Antimicrobial resistance in human isolates of Campylobacter jejuni, 2015.* The Institute of Environmental Science and Research Limited, Porirua, New Zealand, 2015.
- Wu Z, Sahin O, Shen Z, Liu P, Miller WG, Zhang Q.** Multi-omics Approaches to Deciphering a Hyper-virulent Strain of *Campylobacter jejuni*. *Genome Biol. Evol.* 5, 2217–2230, 2013
- Ypma RJF, Bataille AMA, Stegeman A, Koch G, Wallinga J, van Ballegooijen WM.** Unravelling transmission trees of infectious diseases by combining genetic and epidemiological data. *Proc. Biol. Sci.* 279, 444–450, 2012
- Zhang J, Halkilahti J, Hänninen M-L, Rossi M.** Refinement of whole-genome multilocus sequence typing analysis by addressing gene paralogy. *J. Clin. Microbiol.* 53, 1765–1767, 2015

Zhang J, Williamson DA, da Silva Ander G, Biggs PJ, Bulach D, Carter GP, Howden BP, Midwinter AC, Wilkinson D, French NP. *Genomic analysis of the recently emerged antimicrobial resistant Campylobacter jejuni ST-6964 in New Zealand.* In: Characterization of *C. jejuni* and *C. coli* broiler isolates by whole genome sequencing. Pp 152-152. 19th International Workshop on Campylobacter, Helicobacter and Related Organisms, Nantes, France, 2017

**Assessing the impact of host contact network
structure on the phylogeny of *Campylobacter jejuni*
in commercial broiler production systems: a
simulation modelling approach**

Sabrina S. Greening, Ji Zhang, Nigel P. French, and M. Carolyn Gates,

6.1. Abstract

AIMS: Phylogenetic analysis using pathogen sequence data has become an increasingly popular tool for making epidemiological inferences. However, the current lack of robust empirical data on transmission events, combined with the limited understanding of how host population contact structures influence pathogen phylogenies, has made it difficult to validate some of the existing analytical methods. Using *Campylobacter jejuni* transmission in the New Zealand poultry industry as a case example, this study developed a novel disease simulation model to explore the use of pathogen phylogenies in determining (i) who-infected-whom in model systems that consider only discrete network contacts or indeterminate local contacts, and (ii) the relative contribution of different pathways in model systems that consider multiple contact types.

METHODS: Point-mutation within the genome sequence of *C. jejuni* ST-6964 was simulated across 600 outbreaks within the New Zealand commercial broiler industry including 156 farms with a total of 715 poultry sheds. The underlying metapopulation network simulation model incorporated within-farm spread through a stochastic Maternally Immune-Susceptible-Latent-Infectious (MSEI) compartmental model and between-farm spread through both local contacts (determined by the geographical proximity between farms), and network contacts (determined by the reported on- and off-farm movements of transporting vehicles). Three transmission scenarios were explored with 200 simulation replicates with only local contacts, 200 with only network contacts, and 200 with both local and network contacts. All infected poultry sheds were sampled at the end of each simulation replicate and a maximum likelihood phylogenetic tree was constructed based on the number of single nucleotide polymorphisms between the simulated *C. jejuni* sequences. Each phylogeny was then used to infer a single transmission tree which could be compared with the recorded contacts to determine the number of transmissions which had been correctly inferred using the phylogeny. Lastly, a non-parametric Mantel test was completed to explore if the pairwise genetic distance

between the simulated *C. jejuni* sequences could be used to determine the relative contribution of each of the contacts pathways.

RESULTS: Overall, 49.5% (99/200) of the simulations including both local and network contacts resulted in an outbreak that persisted through to the end of the 365-day simulation period. On average, there was 4,625 new shed-level infections across the 715 sheds during the course of a single outbreak, of which 65.0% were a result of a local contact, 30.9% from a network contact, and 4.0% from a neighbouring shed. Among the 200 simulations that modelled the spread of *C. jejuni* through network contacts only, 30.5% (61/200) persisted through end of simulation period with an average of 3,352 new infections during the course of a single outbreak, whilst among the 200 simulations that modelled the spread of *C. jejuni* through local contacts only, 48.0% (96/200) persisted through end of simulation period with an average of 3,959 new infections. On average only 3.0% of the simulated infectious contacts were correctly identified in the inferred transmission trees of which on average 67.6% were local contacts, and 32.4% were network contacts. Results from the non-parametric Mantel test found a weak correlation between the genetic pairwise distance and all other model matrices with little difference seen in the correlation coefficients between the different model systems.

CONCLUSION: This study highlights the importance of combining both network and local contacts in phylogenetic analyses when trying to determine who-infected-whom and mechanisms of spread. Careful consideration of the limitations in pathogen sequence data must also be taken into account, such as the sampling time frame and missing hosts, if pathogen phylogenies are to be used to make any epidemiological inferences.

KEYWORDS: *Campylobacter jejuni*, Phylodynamics, Contact networks, Sequence simulation, Infectious disease modelling

6.2. Introduction

Campylobacter jejuni is one of the leading causes of bacterial foodborne gastroenteritis worldwide (Kaakoush *et al.* 2015) with a large proportion of human campylobacteriosis cases associated with the consumption and handling of poultry meat (Allos, 2001; Humphrey *et al.* 2007). Due to this risk, many countries with commercial poultry production systems have introduced a number of regulatory on-farm prevention measures to reduce the risk to public health, with a clear association seen between high on-farm biosecurity levels and the absence of *C. jejuni* in poultry (Jacobs-Reitsma *et al.* 2008; Lin, 2009). Nevertheless, the effectiveness of individual biosecurity measures in preventing the onward transmission of *C. jejuni* is notoriously difficult to assess due to the pathogen's complex disease dynamics and multiple mechanisms through which it can spread both within and between poultry farms (Newell *et al.* 2011). Further understanding of the risk factors and transmission dynamics of *C. jejuni* in poultry may allow the development of more targeted pre-processing interventions aimed at limiting human exposure to *C. jejuni* (Wagenaar *et al.* 2008).

Previous studies investigating the risk factors and transmission dynamics of *C. jejuni* in poultry have almost exclusively relied on questionnaire-based approaches and risk assessment models with many contradictory results reported between studies (Newell and Fearnley, 2003; Conlan *et al.* 2007). Despite these discrepancies, there are a number of horizontal transmission mechanisms that are thought to be significant risk factors for infection in a commercial broiler flock including the transportation of goods such as feed, litter, and water that are essential for production (Mills and Philips, 2003; Hutchison *et al.* 2004; Newell *et al.* 2010), the movement of personnel including maintenance staff, veterinarians, and catching crews (Slader *et al.* 2002; Ridley *et al.* 2011), and unwanted contacts with wild animals such as birds, rodents, and insects (Craven *et al.* 2000; Ekdahl *et al.* 2005; Nichlos, 2005; Couins *et al.* 2019). With this number of potential transmission routes, it is often difficult in an outbreak situation to know which specific pathways have

contributed to the spread of the disease using contact tracing data. Consequently, a growing number of studies are undertaking phylogenetic analyses using pathogen sequence data isolated from infected farms to make inferences about who-infected-whom (Cottam *et al.* 2008; Jombart *et al.* 2014; de Maio *et al.* 2016), identify risk factors associated with transmissions (Gardy *et al.* 2011; Hayama *et al.* 2019) and determine the transmission mechanisms responsible for the spread of a disease (Spada *et al.* 2004; Ypma *et al.* 2013). These inferences often rely on determining the genetic relatedness between the sampled pathogen isolates based on approaches that have historically been used in other research disciplines such as evolutionary biology and population genetics. Nevertheless, many of the analytical methods integrating pathogen sequence data into epidemiological investigations are still in their infancy and work on only a limited understanding of the complex relationship between disease transmission dynamics and pathogen phylogenies.

To provide insight into this relationship, many studies have begun to incorporate contact network data into phylodynamic analyses to investigate what pathogen phylogenies can reveal about the underlying host population structure (Leventhal *et al.* 2012; Robinson *et al.* 2013; Vasylyeva *et al.* 2016). Contact networks are mathematical and graphical representations of the potential contacts between individuals in a population including features such as the nature, strength, and frequency of the contact which is known to influence the likelihood of a disease spreading via that contact (de Cao *et al.* 2014; Yin *et al.* 2017). However, many contact networks, particularly those for wildlife populations, are difficult to study empirically (Craft, 2015) and present numerous challenges when trying to model disease dynamics using pathogen sequence data. For this reason, some studies have split model systems into discrete network contacts which are more easily defined and local contacts which are more difficult to characterise (Gates *et al.* 2013; Rossi *et al.* 2017; Qi *et al.* 2019). For example, in a study by Firestone and colleagues (2011) social network analysis was used to characterise the importance of both spatial location

and contact network position in the 2007 equine influenza epidemic in Australia. Spatiotemporal models showed that the underlying contact network was essential to describe the pattern of spread whilst models without any reference to the underlying contact network structure were imprecise; particularly in the early phases of the epidemic where it was thought that dissemination was dominated by the movement of infected horses between spatially clustered premises. This study highlighted the importance of linking the spatial and network analyses of disease outbreaks in order to improve the validity of disease transmission inferences and provide insight into both the sequence and scale of spatial spread; helping to pinpoint which pathways may need to be targeted for control (Lessler *et al.* 2016). However, despite the success of these approaches in exploring the relative contribution of different transmission pathways, very few have combined split model systems into a phylogenetic analysis and looked at the ability of phylogenies to capture these dynamics. In fact, until recently there has been very little attempt to validate any of the pathogen phylodynamic models currently in the literature, although a limited number of studies can be found in which the performance of different models have been compared (Klinkenberg *et al.* 2017; Bloomfield *et al.* 2019; Firestone *et al.* 2019).

In this study, the spread of *C. jejuni* through the New Zealand commercial poultry industry was simulated using a metapopulation disease transmission model as a theoretical example to evaluate the validity and performance of the methodology used in Chapter 5 without any intention to make any accurate inferences about the true transmission dynamics of *C. jejuni* in the industry. In the model, the probability of transmission between two farms was modelled as a function of the geographical proximity of the farms as well as the underlying contact network based on the movement of goods and services across the industry. Pathogen sequence data and mutations were also specifically modelled to allow for the reconstruction of pathogen phylogenies. Using both the network data and the sequence data, the objectives of this study were to assess

the performance of methods that use pathogen genetics to determine (i) who-infected-whom for discrete network contacts, (ii) who-infected-whom for nebulous local contacts, and (iii) the relative contribution of different pathways in model systems that consider disease spread through both network and local contacts.

6.3. Material and Methods

6.3.1. Disease simulation model

A disease simulation model was developed to explore how pathogen phylogenies can be used to characterise the spread of *C. jejuni* through the New Zealand commercial broiler industry. The model consisted of four components: (i) a demographic component describing the location and structure of individual poultry farms based on the results from a previous industry survey, (ii) an on-farm component describing the disease status of birds within each poultry shed on farm and the transmission of *C. jejuni* within and between sheds on the same farm, (iii) a between farm component describing disease transmission between different farms *via* one of two mechanisms; indirect local spread and/or direct network contacts and, (iv) a sequence simulation component describing the nucleotide substitutions occurring over time in each infected shed (Figure 6.1).

6.3.2. Model demographics

The model demographics were constructed using the New Zealand commercial broiler industry as a case example. Farm information including the production system (*i.e.*, all-in-all-out for the entire farm versus multiple different age groups simultaneously present on farm), the total number of poultry on farm (*i.e.*, flock size), the number of poultry sheds on farm, the length of the production cycle, and the downtime period between production cycles (where sheds are typically cleaned and left empty for a period between when one grown flock is depopulated and the next batch of chicks is introduced) was collected through a paper-based survey that was sent to all poultry producers registered with either the Poultry Industry Association of New Zealand (PIANZ) or the Egg

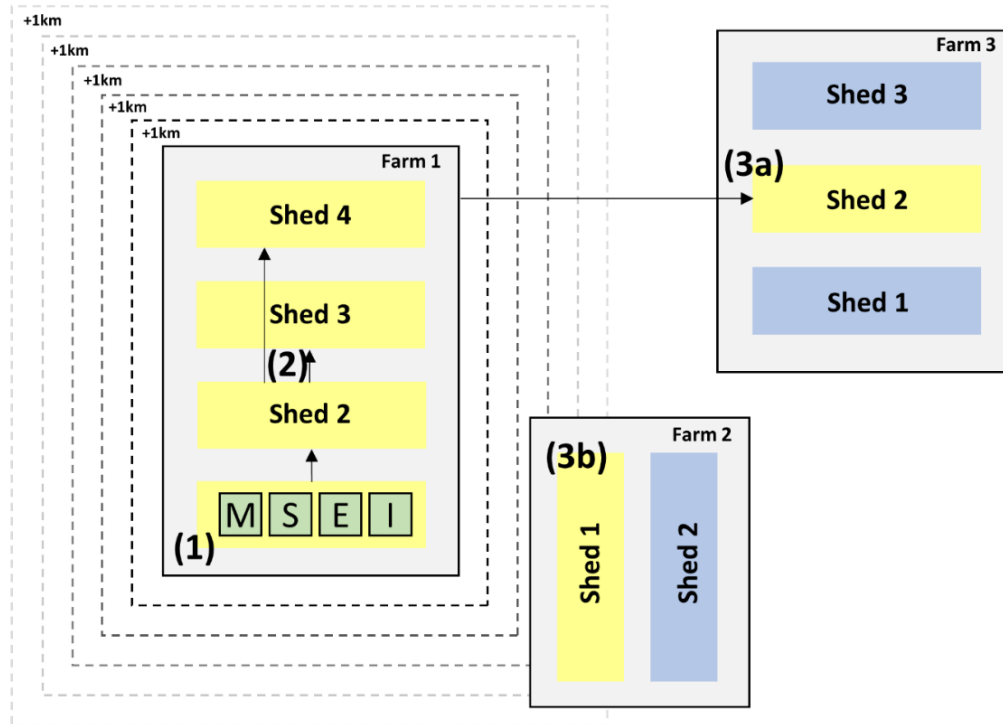


Figure 6.1. Schematic diagram showing an overview of the *Campylobacter* transmission model among three individual farms at time t_n . Each farm has a fixed number of sheds housing a fixed number of birds with the shed colour indicating its infectious status (yellow: infectious and blue: non-infectious). When *Campylobacter* first enters a shed, the within shed prevalence of *Campylobacter* for t_{n+1} is modelled using an MSEI compartment model (1) with the number of infected birds recorded in each time step. The within shed prevalence is used to determine the likelihood of transmission between sheds in the same flock (2) indicated by the vertical arrows in farm 1 where transmission has occurred from shed 1 to shed 2 and from shed 2 to sheds 3 and 4. Transmission between flocks (3) can occur via two mechanism (3a) indicates spread from farm 1 to farm 3 due to a direct contact in the network showing the movement of transporting vehicles (horizontal arrow) whilst (3b) indicates spread from farm 1 to farm 2 due the spatial proximity of the production premises (dashed boxes).

Producers Federation (EPF) as of June 2016. This database was believed to capture the majority of commercial broiler enterprises since PIANZ membership represents over 99% of the country's chicken meat producers. The study was judged to be low risk thorough peer evaluation and consequently was not formally reviewed by any of the University's Human Ethics Committees. Chapter 3 provides a full description of the survey design, implementation, and results, with a copy of the complete survey questionnaire provided in Appendix A. For the purpose of this study, only survey responses from commercial broiler enterprises (defined as farms with > 80% of poultry on-site classified as broilers) were used as they are considered the primary risk factor for *C. jejuni* entering the human food chain (Humphrey *et al.* 2007).

If farms did not respond to the postal survey, the farm demographics (*i.e.*, flock size, number of sheds, production cycle length and downtime period), were inferred by randomly sampling from a kernel density estimate (KDE) derived from a histogram produced using the survey responses. The samples were selected using the R package *stats* (R Core Team., 2018) so that the resulting KDE, derived from the inferred estimates, reproduced the bandwidth and mean from the kernel produced using the survey responses (*i.e.*, the kernel from which the estimates were sampled). Each shed was assigned a random day in the production cycle or downtime period drawn from a uniform distribution to begin the simulations. Sheds on farms with an all-in-all-out system all started the simulations on the same day whereas farms with continuous flow production systems had sheds at different stages of the production cycle. Once set, the day in production on which a shed started each simulation was kept constant over all model iterations.

After initialisation, the farm demographics were kept constant over all model iterations with subsequent management processes represented as deterministic discrete-time events. For example, on the first day in a production cycle on an all-in-all-out farm, the entire farm would be populated with the number of 1-day-old-chicks indicated by their

total flock size with the chicks distributed evenly across all of the sheds present on farm. For a farm with multiple age groups present, sheds were repopulated using the same process, but on different days in the simulation depending on when the individual shed was designated to begin its production cycle. When sheds were repopulated on the first day of the production cycle, it was assumed that chicks were both disease-free and immune to colonisation with *C. jejuni*, regardless of the farm disease status. This assumption was made given strong evidence that the prevalence of *C. jejuni* within a commercial flock is strongly age-dependent and birds less than two- to three-weeks old are rarely colonized naturally (Stern *et al.* 2001; Goddard *et al.* 2013).

Reasons for this lag phase are unclear, but evidence points towards the effect of maternally derived antibodies and age-related differences in the intestinal environment and natural gut flora of birds (van der Wielen *et al.* 2000; Cawthraw and Newell, 2010). However, this naturally immunity decreases over time with most birds shown to become susceptible to colonisation with *C. jejuni* within two- to three-weeks (Stern *et al.* 2000). Immunity to other poultry pathogens that are also known to affect the morbidity and mortality in young chicks, such as *Escherichia coli* (Sawah *et al.* 2018) and *Clostridium perfringens* (Cooper *et al.* 2013), was not accounted for, and the effects of colonisation with these pathogens was reflected in the background mortality rate that was present throughout the duration of the model; such that chicks immune to *C. jejuni* had a higher mortality rate than any other disease status (*i.e.*, susceptible, exposed or infected) to reflect the peak in mortality usually seen 3 to 4 days after placement (Tabler *et al.* 2004; Yassin *et al.* 2009). The model was updated on a daily basis until the end of the production cycle was reached. At this point, either the entire farm was depopulated in all-in-all-out systems or the shed was depopulated on farms with multiple age groups, with sheds remaining empty for the length of the downtime period.

6.3.3. On-farm disease dynamics

A deterministic continuous-time MSEI compartmental model was used to simulate disease dynamics within each shed with birds being classified into one of four mutually exclusive disease statuses: maternally immune (M), susceptible (S), exposed (E), or infectious (I) (Figure 6.2). As each shed started a simulation on a predetermined day in their production cycle, it was assumed that all birds started as susceptible (S) unless the day selected was either within the first 14 days of the farms production cycle, in which case all birds started the simulation in the maternally immune group (M), or the shed started the simulation in their downtime period in which case no birds were present. Throughout the duration of each simulation run, when sheds were repopulated all individuals entered the immune group (M). Birds transitioned into the susceptible group (S), in which they could be infected, at a constant rate (γ) that ensured all individuals would come susceptible after the first 21 days, reflecting the two- to three-week lag phase reported in experimental studies (Stern *et al.* 2000). If a farm became infected during its production cycle, a random shed on the farm was selected, ensuring that the shed contained susceptible birds whilst not already having been infected. One susceptible bird in the selected shed was then moved to the exposed group (E) after which, each susceptible individual in the infected shed received the force of infection (λ) at a rate that $= \lambda(I)$.

Once exposed, birds transitioned into the infectious group (I) at a constant rate (δ) reflecting a latency period of 1.5 days in which exposed birds remained infected but not yet infectious and so therefore did not contribute to the force of infection. Birds in the infected group (I) were assumed to be persistently infected for the remaining duration of the production cycle. vertical lines represent birds leaving the compartment due a background mortality rate (α). In the model all 1-day-olds chicks start immune (M) with maternal antibodies decaying at a constant rate (γ) ensuring that all chicks become susceptible (S) within 2-3 weeks. If infected one susceptible bird was moved to the

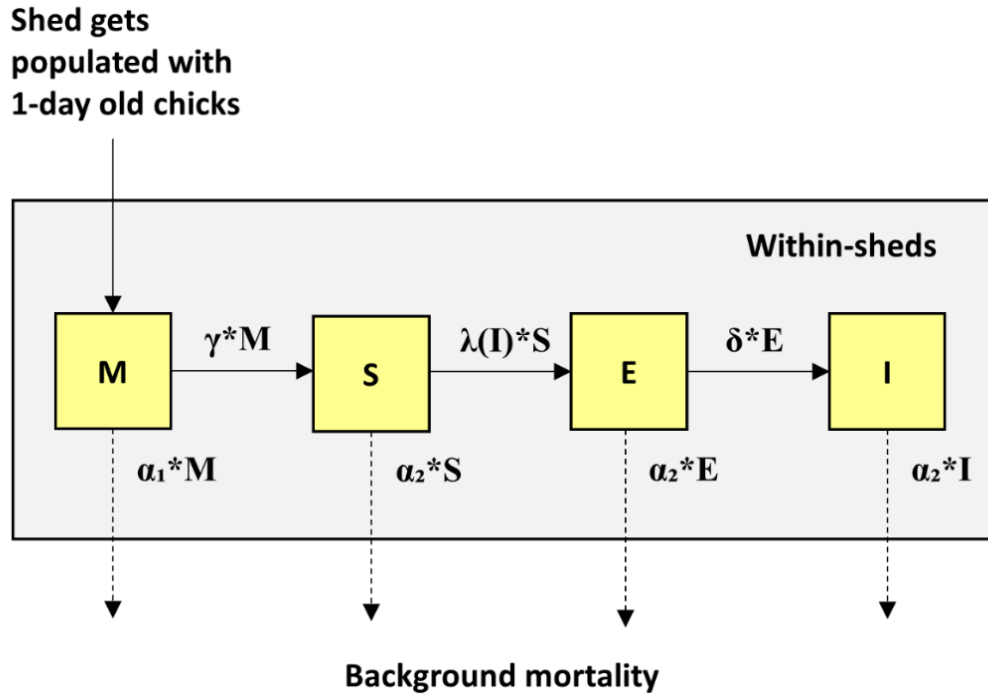


Figure 6.2. Schematic showing the MSEI deterministic compartmental model where solid horizontal lines indicate the movement of birds between compartments whilst dashed that varied as a function of the number of infectious individuals in the shed; such that, λ varied as a function of the number of infectious individuals in the shed; such that, $\lambda = \lambda(I)$. Many different expressions can be used to describe the force of infection; however, for the purpose of this study, λ was generated by considering the interaction between the transmission rate (β), a product of the contact rate and the probability of infection given a contact, and the proportion of infectious individuals (*i.e.*, the disease prevalence: I/N); such that, $\lambda(I) = \beta (I/N)$.

exposed group (E) after which, each susceptible individual in the infected shed received the force of infection (λ) at a rate. Throughout all the simulations the β value was kept constant with an initial value selected to reproduce the dynamics shown in a previous study by van Gerwe and colleagues (2009) in which colonized flocks consisting of 20,000 broiler birds were shown to reach a within-flock prevalence of more than 95% within the first 4 to 8 days after colonization of the first broiler (van Gerwe *et al.* 2009). Once exposed, birds transitioned into the infectious group (I) at a constant rate (δ) reflecting a

mean latency period of 1.5 days (van Gerwe *et al.* 2005) in which exposed birds remained infected but not yet infectious, and therefore did not contribute to the force of infection. Birds in the infected group (I) were assumed to be persistently infected for the remaining duration of the production cycle (Neves *et al.* 2019) giving rise to ordinary differential Equations 1-4, where $\lambda(I) = \beta(I/N)$.

$$\frac{dM}{dt} = \mu N - \gamma M - \alpha_1 M \quad (\text{Equation 1})$$

$$\frac{dS}{dt} = \gamma M - \lambda(I)S - \alpha_2 S \quad (\text{Equation 2})$$

$$\frac{dE}{dt} = \lambda(I)S - \delta E - \alpha_2 E \quad (\text{Equation 3})$$

$$\frac{dI}{dt} = \delta E - \alpha_2 I \quad (\text{Equation 4})$$

Once a shed was infected, the probability of transmission between an infected shed (*i*) and an uninfected shed (*j*) on the same farm was calculated with the probability assumed to scale linearly with the prevalence in the infected shed. If multiple sheds were infected, the daily total probability of transmission was calculated as shown in Equation 5, where *m* is equal to the rate at which the probability increases with the infection prevalence, and $1 - m(I_i/N_i)$ is equal to the probability that transmission will not occur from shed *i* to shed *j*. For the purpose of this study no attempts were made to estimate a value for *m* based on any data in the literature, instead *m* was set to an arbitrary value of 0.05 which remained the same across all the model simulations. This simplifying assumption limits any inference that can be made about the transmission dynamics of *C. jejuni*; however, this was considered beyond the scope of this paper.

$$P_j = 1 - \prod_1^i [1 - m(\frac{I_i}{N_i})] \quad (\text{Equation 5})$$

6.3.4. Between-farm disease dynamics

Transmission between an infected farm and a non-infected farm could occur through two possible pathways including (i) the discrete on- and off-farm movement of commercial vehicles transporting either feed, live birds, hatching eggs, litter, or waste and, (ii) through local contacts including the movements of wild birds, rodents, and other unknown vectors or fomites between neighbouring poultry farms.

6.3.4.1. Spread through network contacts

To determine which farms were connected by the movement of transporting vehicles, three separate contact networks were constructed using information reported in a previous industry survey. Contacts were split to represent common groups of companies that support the poultry industry with one network constructed from all the on- and off-farm movements relating to feed products, another network constructed from all the movements relating to live birds and hatching eggs, including the movement of the catching companies, and the final network constructed from all the movements relating to poultry litter and waste. In each network, the number of nodes was set to equal the number of broiler farms registered in the PIANZ-EPF database ($n = 156$), with edges between nodes representing an indirect connection between farms that had reported using the same transporting company (Figure 6.3).

For any farms that did not respond to the postal survey, missing edges were inferred by grouping non-responders based on two known attributes: farm location and parent company. Non-responders were then randomly assigned a transporting company that was known to service other farms belonging to the same parent company within the same district. For example, a farm belonging to parent company B could only be assigned a transporting company that had been reported in the survey by a farm also belonging to parent company B located in the same district as the non-responder. For each inferred edge, the frequency of movements along that edge also had to be inferred. Values for the

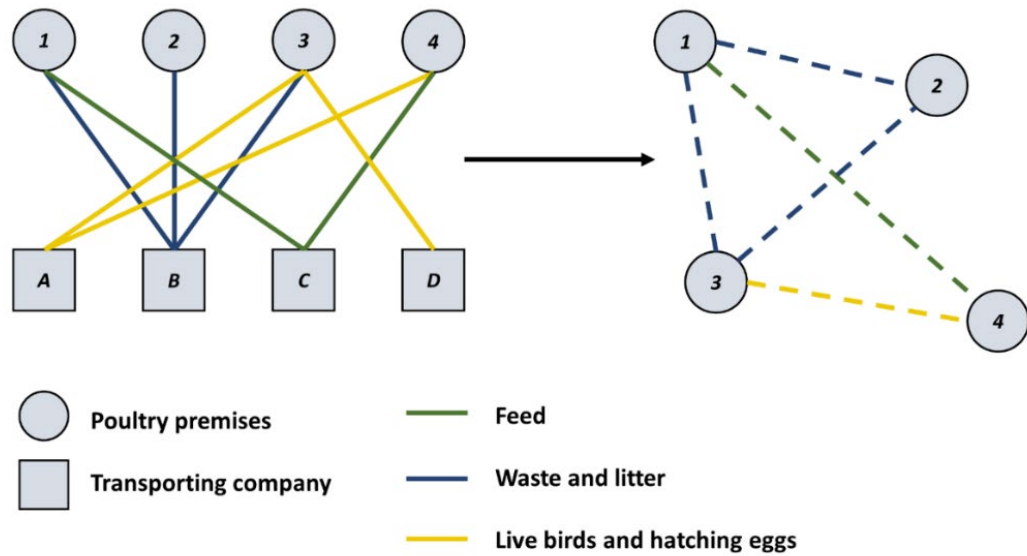


Figure 6.3. A schematic diagram showing a network graph (left) that could have been constructed from survey responses with vertices (V) belonging to different groups; for example, V1 (1, 2, 3 and 4) represent poultry farms whilst V2 (A, B, C and D) represent those transporting companies reported to have delivered goods or services to the connected farm. This graph can be used to construct the network on the right which contains just the vertices from the V1 group (*i.e.*, only the poultry farms that responded to the survey), with edges between vertices that share a common vertex belonging to the V2.group. For example, the green dashed edge connects farm 1 and farm 4 in the right-hand graph due to the movement of waste and litter shown in the left-hand graph by transporting company C.

frequency were estimated similar to the missing farm demographics with estimates randomly sampled from a KDE derived from the histogram produced using the survey responses so that the resulting KDE reproduced the bandwidth and mean from the kernel they were sampled from. Edge frequency was used to determine the number of days between contacts along an edge in the model. For example, a contact frequency of 10, indicated that there was 10 days between two consecutive contacts along that edge. In addition to edge frequency, a number of network statistic were used to describe the structure of each of the inferred networks including (i) degree, (ii) betweenness, (iii)

density, (iv) average path length, (v) clustering coefficient, (vi) fragmentation and, (vii) diameter. All network statistics were calculated using the R package *igraph* (Csardi and Nepusz, 2006) with a further description of each measure provided in Table 6.1. Given a contact in any of the networks, the probability of transmission between an uninfected farm (j) and an infected farm (i) was calculated with the probability assumed to scale linearly with the prevalence of *C. jejuni* infected birds on the infected farm. If a farm had multiple contacts in the same network on the same day, the total probability of transmission was given as shown in Equation 6.

$$P_j = 1 - \prod_i [1 - (\frac{I_i}{N_i} \phi e^{-\alpha d_{ij}})] \quad (\text{Equation 6})$$

Table 6.1. Network analysis glossary of terms used to describe the contact network graphs representing the movement of goods and services between commercial poultry producers in New Zealand.

Network measure	Definition
<i>Total degree</i>	The sum of the in- and out-degree, whereby the in-degree is the number of individual movements onto a farm and the out-degree is the number of individual movements onto a farm
<i>Betweenness</i>	The frequency a farm is found on the shortest path between any other two pair of farms in the network
<i>Density</i>	The proportion of all possible links between farms in the network that are present
<i>Average path length</i>	The average shortest path between any pair of farms in the network averaged over all pairs of farms
<i>Clustering coefficient</i>	The proportion of neighbouring farms in direct contact with a farm that are also connected to each other.
<i>Fragmentation</i>	The proportion of farm pairs for which a path does not exist between them.
<i>Diameter</i>	The longest path between any two pair of farms in the network

6.3.4.2. Spread through local contacts

In addition to network contacts, local contacts were considered possible between all farms within 7km (~4.34 miles) of each other. This distance was selected on the basis of the highly pathogenic avian influenza buffer zones recommended by the United States

Department of Agriculture (USDA) which consider farms outside this distance from an infected farm, being unlikely to get infected (USDA, 2017). The Euclidean distance between each farm was calculated using the addresses provided in the PIANZ-EPF database with addresses first having been checked using Google Maps (2017) to make sure they specified a poultry farm (indicated by the presence of poultry sheds) and not the producers' residential address. The probability of transmission between an uninfected farm (j) and an infected farm (i) was then calculated with the probability assumed to scale linearly with the prevalence on the infected farm whilst decreasing exponentially with the Euclidean distance between the two farms (d_{ij}), and if a farm had multiple infected neighbours, the daily total probability of transmission was given as shown in Equation 6, where α is the rate at which the probability declines with distance, \emptyset is equal to the probability of transmission when the d_{ij} is equal to zero, and $1-(I_i/N_i)\emptyset e^{-\alpha d_{ij}}$ is the probability that transmission will not occur. For the purpose of this study no attempts were made to estimate a value for either \emptyset or α based on data in the literature, instead \emptyset was set to an arbitrary value of 0.05 whilst α was set to 1.46, similar to the transmission kernel applied in a previous study by VanderWaal and colleagues (2016), with both \emptyset and α remaining the same across all the model simulations.

6.3.5. Estimating the probability of infection

Once the transmission probability for each pathway had been calculated, the total probability of a new infection occurring on a farm was calculated as shown in Equation 7; taking in to account all the possible routes including between shed transmission (if the farm was already infected), transmission via a transporting vehicle, or transmission via a local contact (Figure 6.4). A random number generator was then used to select any number between zero and one (n) and if the total probability of infection was greater than n , an infection was seeded on the farm. Given that an infection occurred, a Bernoulli trial was then performed to select which pathway would be considered responsible for the infection, followed by an addition trial to select the source farm out of the possible

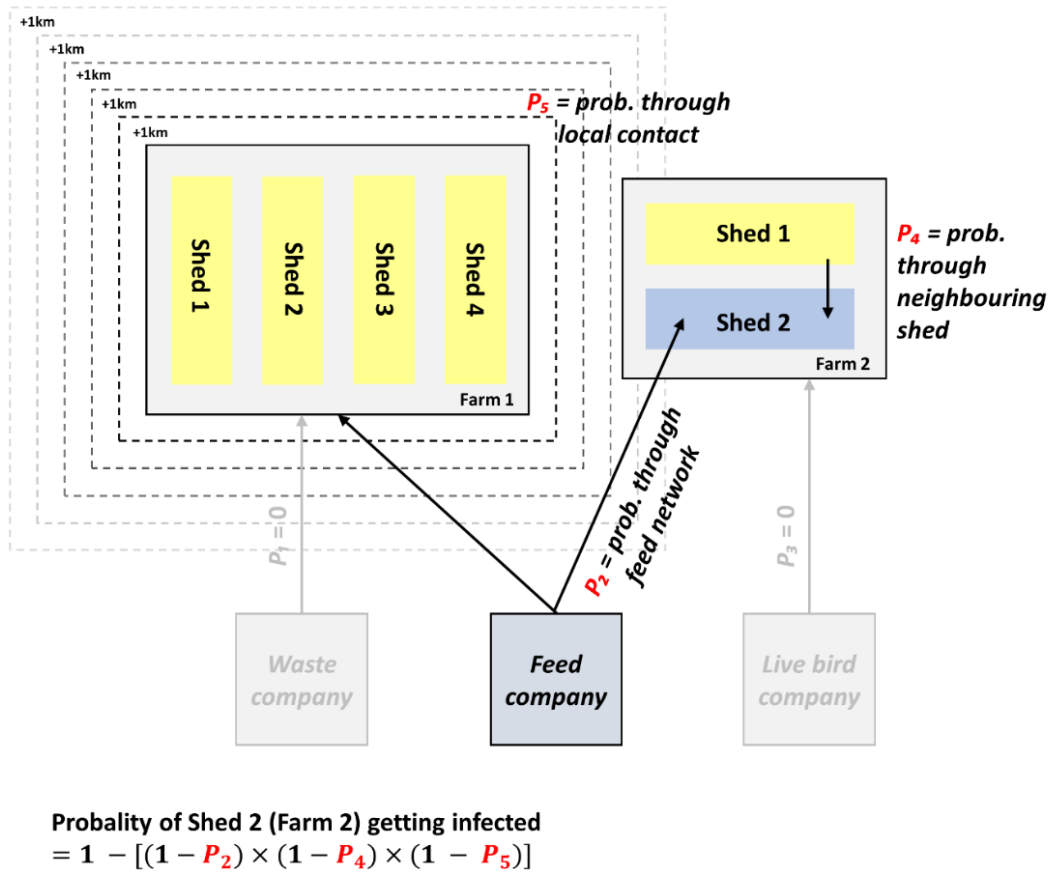


Figure 6.4. A schematic diagram showing how to calculate the total probability (P) of a new infection occurring in an uninfected shed (*i.e.*, Shed 2, Farm 2: shown in blue) from an infected shed (shown in yellow). In this example no infection can result from a movement in either the waste network, or the live bird and hatching egg network as the farms do not share the same transporting companies. Therefore, both P_1 and P_3 are equal to 0. However, there is a probability of Shed 2 becoming infected from Farm 1, in which all sheds are infected, due to either the close proximity of the two farms (indicated by the dashed lines) and the probability of a local contact (P_5), or via a contact in the feed network (P_2), as both Farms 1 and 2 share the same feed company. Shed 2 could also become infected via contact with the other infected shed on the farm (*i.e.*, Shed 1, Farm 2). Therefore, the total probability of Shed 2 becoming infected is: $1 - [(1 - P_1) \times (1 - P_2) \times (1 - P_3)]$.

infected farms in contact with the newly infected farm on the selected pathways (*i.e.*, if an uninfected farm had multiple infected neighbours, the source was determined through a Bernoulli trial using the individual probabilities of each infected neighbouring farm). A shed was then randomly selected from the source farm, ensuring that it had at least one infectious bird, to be recorded as the source shed.

$$P_j = 1 - \prod_1^i (1 - P_i) \quad (\text{Equation 7})$$

6.3.6. Sequence simulation

For the sequence simulation, a single *C. jejuni* ST-6964 whole-genome sequence was selected from the whole-genome sequencing data presented in Chapter 5; that is, the 167 *C. jejuni* isolates that were sampled as an extension of the study reported by the Institute of Environmental Science and Research (ESR) in collaboration with the Poultry Industry of New Zealand (PIANZ) (Muellner *et al.* 2016). For the full details describing sample collection, isolate culture and the genomic analysis of the isoaltes including genome sequencing, assembly and multi-locus sequence typing (MLST) readers are directed to Chapter 5. For the purpose of this study, the single *C. jejuni* ST-6964 sequence was selected based on two quality control features; the number of overlapping reads (*i.e.*, contigs) and the guanine-cytosine content (Gurevich *et al.* 2013) . The selected sequence was then limited to the size of the first node produced from *de novo* assembly: that was, 208,001 nucleotide bases, as a trade-off between computational capacity and biological accuracy. The sequence was used to seed the infection in each model simulation with the sequence remaining the same across all iterations.

Given an infectious contact, the newly infected shed would acquire the sequence from the source shed and a nucleotide substitution process would pursue independently in both infected sheds following the Jukes-Cantor model for nucleotide substitution (Jukes and Cantor, 1969) under an adjusted substitution rate based on the reported rate of 3.0×10^{-7}

per site per year for this *C. jejuni* lineage (French *et al.* 2019). This substitution rate was adjusted, as shown in Equation 8, to provide a daily probability of mutation over the limited number of nucleotide bases in the selected sequence, with the adjusted rate being used to determine the number of potential substitutions occurring in each time step under a binomial distribution: $B(n, p)$, where n is equal to the sequence length and p equal to the adjusted substitution rate. Once infected and a sequence had been recorded for a shed, it was assumed that multiple infections could not occur within the same production cycle meaning that each infected shed only had a single sequence.

$$\text{Adjusted substitution rate} = \frac{[(3.0 \times 10^{-7} \times 1,708,234)/208,001]}{365}$$

Equation 8. Calculation used to adjust the substitution rate provided by French and colleagues (2019) (*i.e.*, 3.0×10^{-7} per site per year) to give a daily substitution rate across only a small length of the genome (*i.e.*, 208,001 bp out of 1,708,234 bp)

6.3.7. Simulations conditions

Each model simulation ran over the course of one year (*i.e.*, 365 days) unless the outbreak died out after the initial seed was set. To seed the infection at the start of each simulation, a single shed was randomly selected. If the selected shed had either all immune birds or no birds (*i.e.*, within the first 14 days of production or in the downtime period), the selection process was repeated until a susceptible shed was chosen, in which case four birds were moved from the susceptible group (S) into the exposed group (E). With each new infection, the infectious pathway, the source farm, the source shed, the day of infection, and the genetic sequence in the source shed at the time of infection was recorded and extracted as a CSV file. This information was used to (i) track who-infected-whom, (ii) determine the contribution of each contact pathway to the number of new infections, and (iii) produce epidemic curves showing the total number of infected farms on each day of the simulation. The genetic sequence was then updated every day to record

any mutations that may have occurred with all genetic sequences at the end of each simulation (*i.e.*, one sequence from all infected sheds after 365 days), exported as a multi-sequence alignment FASTA files for use in the phylogenetic analysis described in section 6.1.8. Altogether 600 model iterations were run, with 200 including both local and network contacts, 200 including just local contacts, and the remaining 200 including just network contacts.

6.3.8. Phylogenetic analysis

In order to assess the ability of a phylogeny to capture the contribution of multiple transmission pathways on the spread of the outbreak, a phylogenetic tree was constructed at the end of each simulation. Maximum Likelihood (ML) phylogenetic trees were constructed using MEGA7 (v7.0.26) (Kumar *et al.* 2016) based on the number of Single Nucleotide Polymorphisms (SNPs) in the multi-sequence alignment and the Jukes-Cantor model (Jukes and Cantor, 1969), with SNPs extracted using the software tool SNP-sites (v3.0.) (Page *et al.* 2016). The pairwise genetic distance between each of the sequences sampled at the end of the simulation was also calculated based on the uncorrected p -distance measure; that is, the proportion of nucleotide sites that differ between each pair of sequences. P -distance matrices were constructed using the R package *ape* (Paradis and Schliep, 2018) with no corrections made for multiple substitutions at the same site. The uncorrected p -distances were used in the statistical analysis described in section 6.1.9.

To infer who-infected-whom from the simulated *C. jejuni* sequences, the computational method for the reconstruction of transmission trees described by Klinkenberg and colleagues (2017) was implemented using the R package *phybreak* (Klinkenberg *et al.* 2017). To reduce the computation time, only 50 (50.5%, 50/99) of the simulations including both network and local contacts were randomly selected for this analysis in which a Markov-Chain Monte Carlo (MCMC) chain was run with a burn-in length of

5000 MCMC cycles followed by 25,000 samples. A final consensus transmission tree was constructed for each of the 50 simulations by finding the maximum parent credibility (mpc) tree as described in Hall and colleagues (2015). This is the tree in the set of posterior samples that has maximum support as measured by the frequency of the infector in the posterior distribution. The consensus transmission tree was then compared with the recorded transmission network (Figure 6.5) to determine (i) the proportion of direct infectious contacts correctly identified (*i.e.*, who-infected-whom), and (ii) the proportion of contacts in the transmission chain correctly identified (*i.e.*, direct infectious contacts that have been incorrectly inferred but are still connected in the same transmission chain).

6.3.9. Statistical analysis

To examine if the phylogeny had captured the contribution of all the different contacts the relationship between the genetic matrix and an additional four individual model distance matrices was examined using a non-parametric Mantel test (Mantel, 1967). Out of the four additional matrices, there was a single geographical distance matrix in which values corresponded to the Euclidean distance between each poultry shed in the model. This matrix was constructed using the R package *geosphere* (Hijmans, 2019) and the farm coordinates that had been extracted from the PIANZ-EPF database. Each value in the matrix was expanded to express the pairwise Euclidean distance (km) between the sheds on each farm with values for sheds on the same farm set to zero. The three remaining model distance matrices represented the shortest path between farms in each of the contact networks, that is, the minimum number of links between two nodes in a network weighted by the frequency of the movement along that edge. To calculate the shortest path, network graphs were constructed using the R package *igraph* (Csardi and Nepusz, 2006) under the force-based algorithm proposed by Fruchterman and Reingold (1991). These matrices were then also expanded to express the pairwise distance between each model shed, however, if a pair of farms were completely unconnected in the network

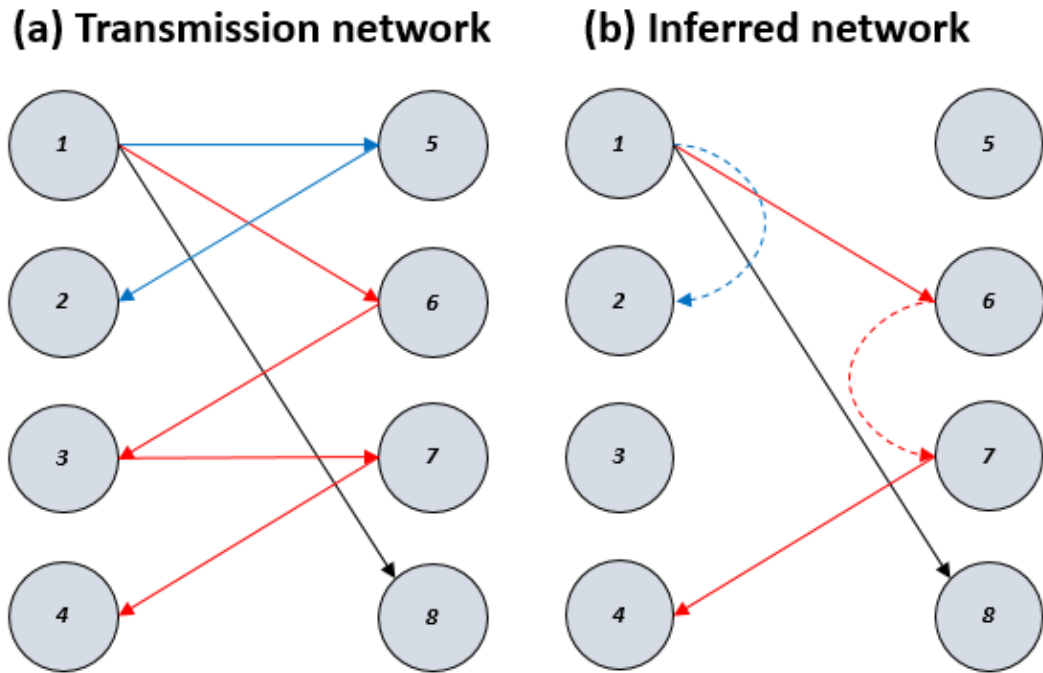


Figure 6.5. (a) A schematic diagram representing a transmission network containing three different transmission chains as indicated by the colour of the lines (blue, red and black). In this example, the primary case is Farm 1, who directly infects Farm 5 (blue chain), Farm 6 (red chain), and Farm 8 (black chain). There is then onward transmission from both Farms 5 and 6 as indicated by the direction of the arrows in each of the chains. (b) shows the corresponding network that has been inferred using pathogen sequence data. In the network the direct transmissions between Farm 1 and Farm 6, Farm 1 and Farm 8, and Farm 7 and Farm 4 have been correctly identified, as indicated by the solid lines. Indirect transmission between Farm 1 and Farm 2 as well as between Farm 6 and Farm 7 has also been correctly identified, as indicated by the dashed line, meaning that these farms are not direct contacts in (a) but are in the same transmission chain. Therefore, the proportion of infectious contacts correctly identified is 42.9% (3/7), whilst proportion of contacts in a transmission chain correctly identified 71.4% (5/7)

graphs, a numeric value of 99 was recorded in the matrix in order to be able to distinguish between those sequences sampled from sheds on unconnected farms and those sequences sampled from sheds on the same farm.

A non-parametric Mantel test (Mantel, 1967) was then implemented between the genetic p -distance matrix and each of the individual model distance matrices using the RELATE routine in PRIMER (v7.0.) (Clarke and Gorley, 2015), with Spearman's rank correlations (ρ) calculated as a measure of matrix correlation. To further examine if any of the phylogenies had captured the contribution of local contacts, tanglegrams were constructed using the R package *dendextend* (Galili, 2015) in which each sequences on the phylogenetic tree was connected *via* an auxiliary line to the corresponding node on a hierarchical dendrogram representing the Euclidean distance between the sheds from which the sequences were sampled from. To optimise each tanglegram, a two-tree crossing minimization technique was used to minimize the number of crossings between the auxiliary lines. Tanglegrams were then used to calculate the cophenetic correlation coefficient between the trees cophenetic distances matrices with values near 0 meaning that the two trees are not statistically similar (Sokal and Rohlf, 1962) and suggesting that the geographical proximity of farms has little influence on the shape of the pathogen phylogeny.

6.4. Results

6.4.1. Industry demographics and contacts

Out of the 156 broiler farms registered in the PIANZ-EPF database, 36.5% (57/156) responded to the industry survey, meaning that the farm demographics for 99 of the farms in the model had to be inferred; overall resulting in a final model with 715 poultry shed and the model demographics summarised in Table 6.2. The average Euclidean distance between a pair of farms was 309.63km (min: 0.001, max: 1049.71, median: 212.95) with only 4.0% (480/12,090) of pairwise distances being 7km or less: the cut-off distance for

local spread to occur between two farms (Figure 6.6). Across the three inferred contacts networks, the waste and litter network had the greatest average path length equal to 2.15 followed by the live bird and hatching eggs network with an average path length of 2.10 and the feed network with an average path length of 1.27. The live birds and hatching eggs network had the greatest number of edges with 9722 edges between the 156 broiler farms (Table 6.3).

Table 6.2. Summary farm demographics across 156 broiler farms registered in the PIANZ-EPF database, of which 63.5% (99/156) have been inferred by randomly sampling from a kernel density estimate derived from a histogram produced using survey responses from 36.5% (57/156) of the broiler farms.

Farm demographic		Number of farms (%)
Housing type	<i>Free-range</i>	28 (17.9)
	<i>Barn</i>	120 (76.9)
	<i>Colony</i>	8 (5.1)
Flow	<i>All-in-all-out</i>	112 (71.8)
	<i>Multiple age groups</i>	44 (28.2)
Mean flock size (min-max)		135,767.7 (31,200-923,000)
Mean number of sheds (min-max)		4.6 (2-12)
Mean production cycle length (days) (min-max)		42.4 (18-56)
Mean downtime period (days) (min-max)		8.9 (0-18)

Table 6.3. Summary network measures for the inferred network graphs representing the yearly on- and off-farms movements relating to feed, live birds and hatching eggs and, poultry waste and litter between all 156 commercial broiler producers in New Zealand registered on the PIANZ-EPF database.

Network measure	Network		
	Feed	Live birds and hatching eggs	Waste and litter
<i>Number of edges</i>	6038	9722	9188
<i>Average total degree</i>	35.87	38.31	33.35
<i>(min-max)</i>	(3-62)	(3-86)	(4-80)
<i>Average betweenness</i>	8.09	148.7	153.80
<i>(min-max)</i>	(0-1050)	(0-4903)	(0-6345)
<i>Density</i>	0.23	0.25	0.22
<i>Average path length</i>	1.27	2.10	2.15
<i>Clustering coefficient</i>	0.99	0.91	0.78
<i>Fragmentation</i>	0.69	0.00	0.10
<i>Diameter</i>	550	353	377

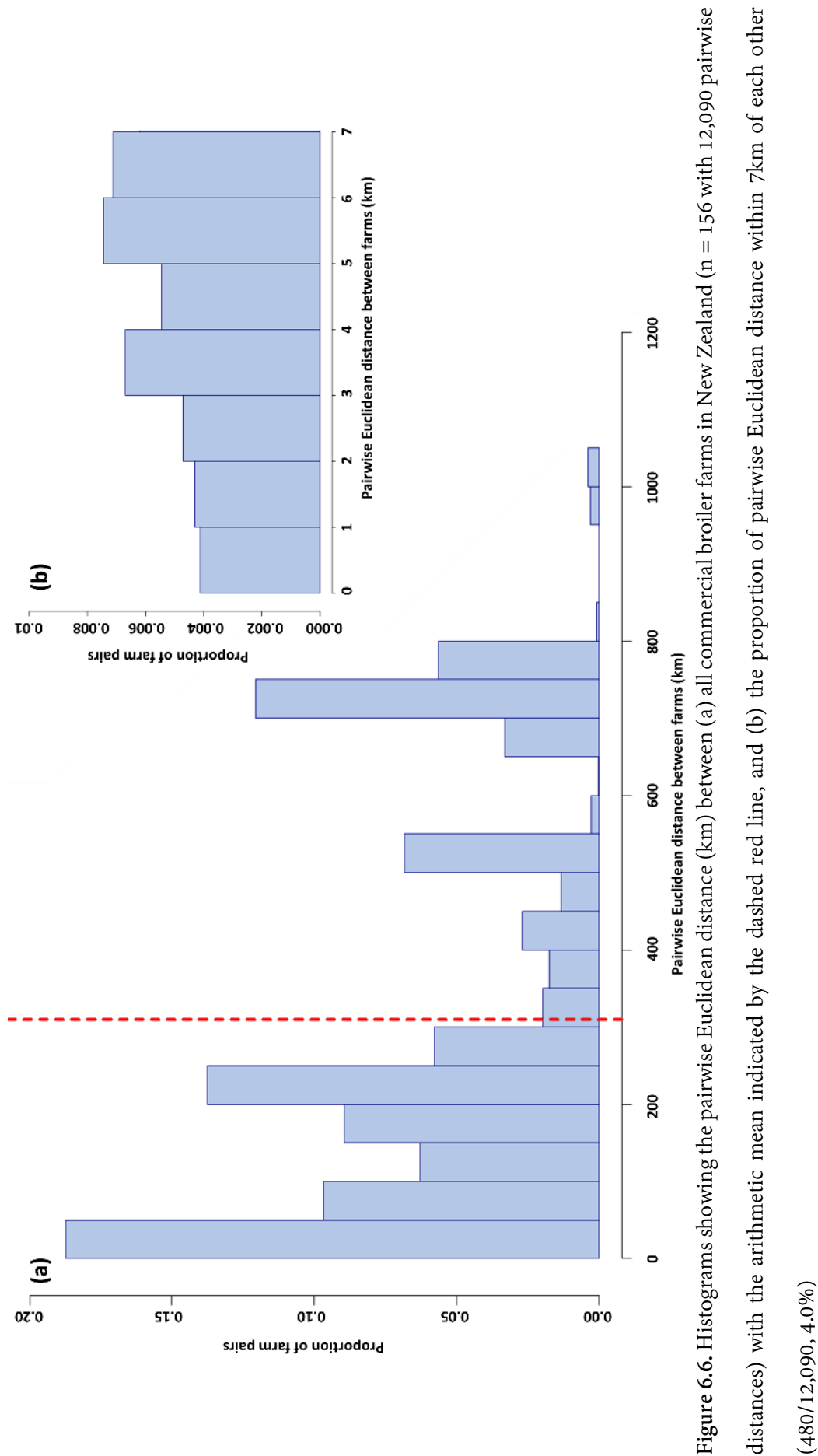


Figure 6.6. Histograms showing the pairwise Euclidean distance (km) between (a) all commercial broiler farms in New Zealand ($n = 156$ with 12,090 pairwise distances) with the arithmetic mean indicated by the dashed red line, and (b) the proportion of pairwise Euclidean distance within 7km of each other (480/12,090, 4.0%)

6.4.2. Disease transmission model

Among the 200 simulations that modelled the combined spread of *C. jejuni* through both network and local contacts, 34.5% (69/200) failed to generate a secondary infection after the initial seed, 16.0% (32/200) died out before the end of simulation period, and 49.5% (99/200) persisted through end of simulation period. Across all the model iterations that resulted in the onward spread of *C. jejuni* and in which the outbreak ran for the full length of the simulation, there was an average of 4,625 (min: 3,918, max: 4,907, median: 4,689) new infections during the course of a single outbreak, of which an average of 65.6% (min: 64.0, max: 67.2, median: 65.6) were a result of a local contact, 18.0% (min: 16.7, max: 19.4, median: 18.0) from a contact in the feed network, 8.0% (min: 7.1, max: 9.2, median: 8.0) from a contact in the live bird network, 4.8% (min: 4.1, max: 5.8, median: 4.8) from a contact in the waste network and, 3.6% (min: 2.9, max: 4.4, median: 3.6) from a neighbouring shed. On average 58.3% (91/156) of the farms were infected on day 90 in comparison to 85.9% (134/156) on day 180, 82.7% (129/156) on day 270, and 85.9% (134/156) on the last day of the simulation, with the complete epidemic curves presented in Figure 6.7.

Across the 200 simulations that modelled the spread of *C. jejuni* through only network contacts, 45.0% (90/200) failed to generate a secondary infection after the initial seed, 24.5% (49/200) died out before the end of simulation period, and 30.5% (61/200) persisted through end of simulation period. Across all the model iterations that resulted in the onward spread of *C. jejuni*, and in which the outbreak ran for the full length of the simulation, there was an average of 3,352 (min: 2,556, max: 3,997, median: 3,356) new infections during the course of a single outbreak, of which on average 52.0% (min: 50.0, max: 53.8, median: 52.1) were a result of a contact in the feed network, 20.0% (min: 19.0, max: 21.6, median: 20.1) from a contact in the live bird network, 14.4% (min: 13.2, max: 15.8, median: 14.4) from a contact in the waste network, and 14.4% (min: 13.2, max: 15.8, median: 14.4) from a neighbouring shed. On average 15.4% (24/156) of the farms were

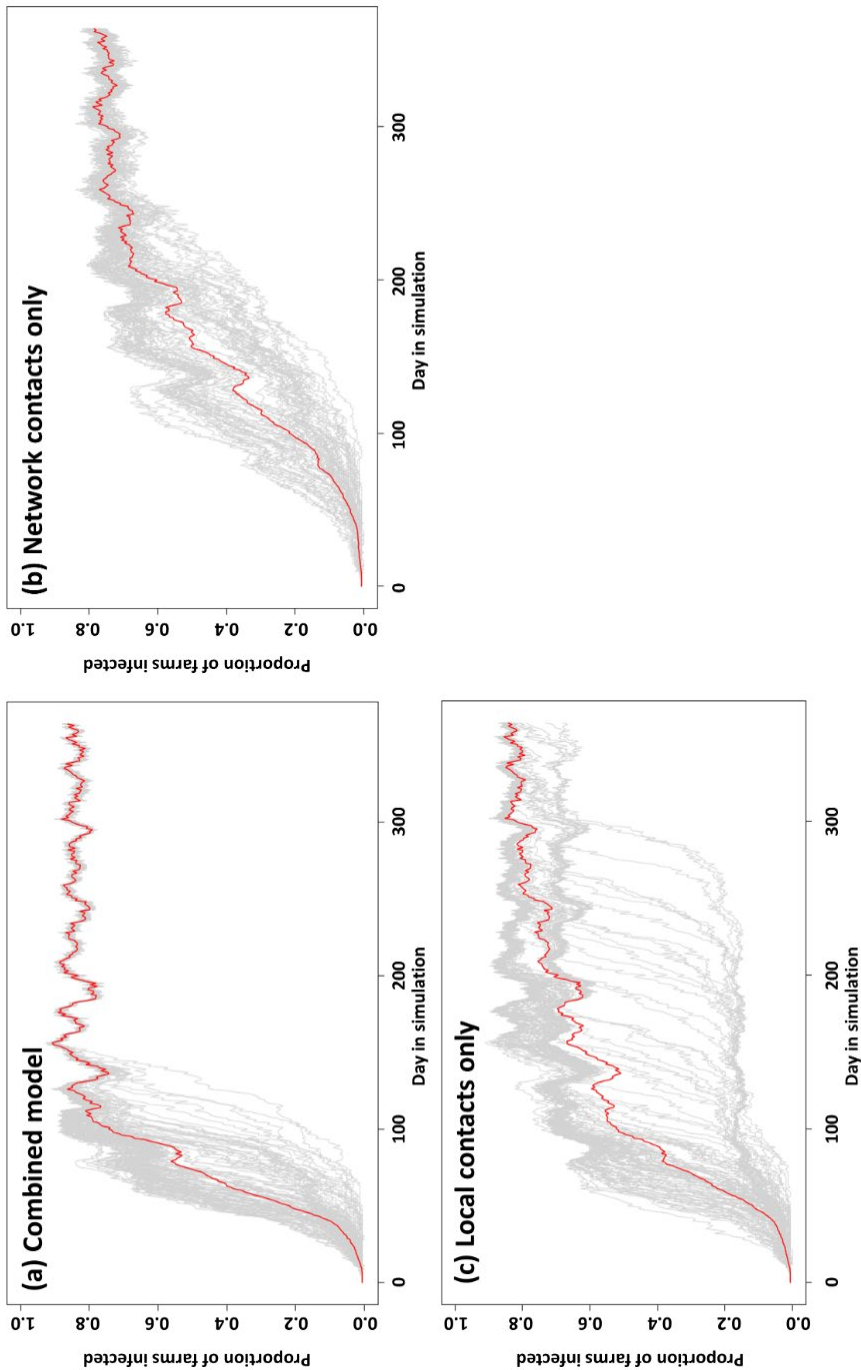


Figure 6.7. Epidemic curves showing the proportion of all broiler farms ($n = 156$) that got infected on each day across all the simulations which the outbreak ran for the full 365 days including (a) 99 simulations with both local and network contacts, (b) 61 simulations with only network contacts, and (c) 96 simulations with only local contacts. The red line on each graph indicates the mean across all the model simulations.

infected on day 90 in comparison to 56.4% (88/156) on day 180, 73.1% (114/156) on day 270, and 78.8% (123/156) on the last day of the simulation (Figure 6.7).

Among the 200 simulations that modelled the spread of *C. jejuni* through only local contacts, 38.0% (76/200) failed to generate a secondary infection after the initial seed, 14.0% (28/200) died out before the end of simulation period, and 48.0% (96/200) persisted through end of simulation period. Across all the model iterations that resulted in the onward spread of *C. jejuni*, and in which the outbreak ran for the full length of the simulation, there was an average of 3,959 (min: 2,095, max: 4,677, median: 4,162) new infections during the course of a single outbreak, of which 94.5% (min: 92.9, max: 95.7, median: 94.5) were a result of a local contact, and 5.5% (min: 4.3, max: 7.1, median: 5.5) from a neighbouring shed. On average 41.7% (65/156) of the farms were infected on day 90 in comparison to 67.9% (106/156) on day 180, 77.6% (121/156) on day 270, and 84.0% (131/156) on the last day of the simulation (Figure 6.7).

6.4.3. Phylogenetic and statistical analyses

Results from the non-parametric Mantel test found a weak correlation between the p -distance matrix (indicating the genetic pairwise distance between sample isolates) and the Euclidean distance matrix (indicating the geographical distance between infected sheds) with an average cophenetic correlation coefficient equal to 0.017 (min: -0.077, max: 0.229, median: 0.010) across simulations including both network and local contacts, 0.028 (min: -0.072, max: 0.277, median: 0.014) across simulations including only network contacts, and 0.033 (min: -0.086, max: 0.474, median: 0.009) across simulations including only local contacts (Figure 6.8). The average Spearman ρ values between the phylogenetic p -distance matrix and the different network model matrices also showed very little variation across the different model systems, even in comparison to the model simulations that did not include any network contacts (Figure 6.9). For example, the average Spearman ρ

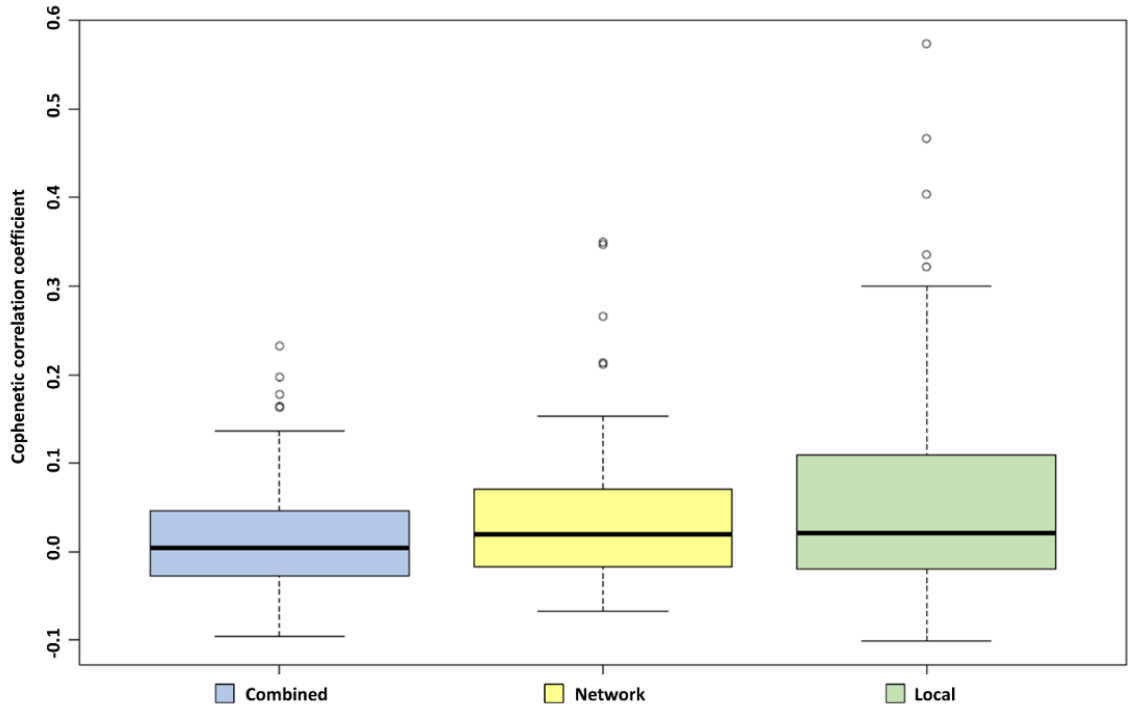


Figure 6.8. Boxplot showing the cophenetic correlation coefficient over 99 model simulations including both network and local contacts (blue), 61 model simulations including only network contacts (yellow), and 96 model simulations including only local contacts (green) with values in the graph representing the minimum, maximum, median, first quartile and third quartile in the data.

values between the phylogenetic p -distances and the distance in the feed network was 0.014 (min: -0.051, max: 0.116, median: 0.007) across all simulations that included both network and local contacts, 0.021 (min: -0.036, max: 0.186, median: 0.010) across all simulations that only included network contacts, and 0.018 (min: -0.051, max: 0.253, median: 0.008) across all simulations that only included local contacts. The cophenetic correlation coefficient also showed very little variation between the model systems with an average value of 0.018 (min: -0.095, max: 0.232, median: 0.004) across all simulations including both network and local contacts, 0.042 (min: -0.067, max: 0.350, median: 0.020) across all simulations including only network contacts, and 0.062 (min: -0.101, max: 0.574, median: 0.021) across all simulations including only local contacts (Figure 6.9).

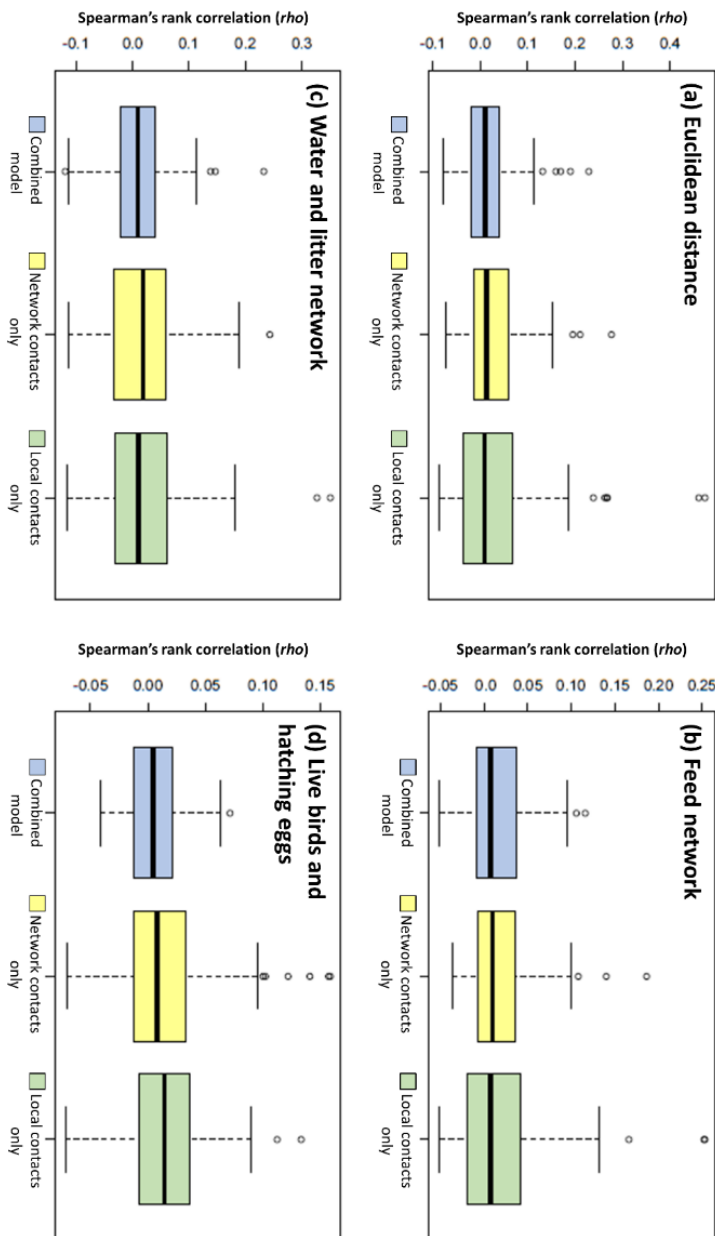


Figure 6.9. Boxplot showing the Spearman's rank matrix correlation (ρ) between the model matrices for (a) the Euclidean distance between poultry farms, (b) the feed network, (c) the waste and litter network and, (d) the live bird and hatching eggs network and the genetic p -distance matrix between all simulated *C. jejuni* sequences sampled on the 365th day of each model simulation including 99 simulations with both network and local contacts (blue), 61 simulations with only network contacts (yellow), and 96 simulations with only local contacts (green). Values in the graph represent the minimum, maximum, median, first quartile and third quartile in the data.

When comparing the consensus transmission trees, inferred from each phylogeny, with the recorded transmission networks from across 50 of the simulations including both network and local contacts, on average 3.0% (min: 1.7, max: 5.3, median: 2.9) of direct infectious contacts were identified in the consensus tree whilst 47.4% (min: 41.6, max: 50.3, median: 47.9) of infectious contacts in the transmission chain were identified. Out of those direct contacts correctly inferred on average 61.3% (min: 30.0, max: 84.6, median: 63.1) were a local contact, 17.5% (min: 0.0, max: 50.0, median: 16.7) were a contact in the feed network, 9.7% (min: 0.0, max: 31.3, median: 10.0) were a contact in the live bird network, 5.2% (min: 0.0, max: 25.0, median: 5.4) were a contact in the waste network, and 6.4% (min: 0.0, max: 30.8, median: 5.1) were a contact with a neighbouring shed on the same farm.

6.5. Discussion

To the best of our knowledge, this is the first disease transmission model that has been constructed within a commercial poultry industry that incorporates pathogen genome sequence data and, despite the limited availability of demographic and contact network data, this model presents a framework for assessing the use of pathogen phylogenies in determining important transmission dynamics such as who-infected-whom. However, in order to make any inferences about the spread of *C. jejuni* as the model pathogen, more work is needed to improve the estimates for the model parameters including the transmission rate. Nevertheless, the variation in transmission patterns across the model systems highlights the importance of combining both network and local contacts in future models to ensure meaningful inferences can be made about disease outbreak dynamics.

Despite of the variation in transmission patterns across the different model systems, very little distinction was identified using the Mantel test; with no association found between the pairwise genetic distance of simulated sequences and any of the model matrices. This

result does not concur with the findings presented in Chapter 5 or the general hypothesis that a disease spreading predominantly through a single contact, for example local contacts, would result in a strong correlation between the phylogeny, indicating the genetic distance between isolates, and the corresponding distance matrix (*i.e.*, the Euclidean distance). However, it is important to consider that the lack of correlation in the modelled dataset in comparison to the data presented in Chapter 5 may be due to some of the limiting assumptions in the model structure. For example, when considering the spread through local contacts a fixed boundary was used so that an infected farm was only able to infect those farms within a 7km radius. Not only is this unlikely to be the case in reality but it also limits local spread to only a small proportion of farm pairs as shown in Figure 6.6. Therefore, the Euclidean distance matrix, which was used in the Mantel test and accounts for all the pairwise distances between farms including those far greater than 7km, may not accurately reflect local contacts due to the fixed limit used in the model. This may explain why the correlation between the genetic distance and the Euclidean distance is lower than expected as indicated by the *rho* values (*i.e.*, the Mantel test); although for the cophenetic correlation coefficients that were calculated from the tanglegrams, no significant correlation was found in either the modelled dataset or the data presented in Chapter 5. This highlights the importance of assessing the performance of analytical methods that are currently being developed and considering the effects of integrating epidemiological data into a phylogenetic framework (or vice versa). For instance, for the cophenetic correlation coefficients it is important to consider the effect of comparing a dendrogram with a phylogenetic tree whose structures are not quite the same; with the former including the lengths of the branches and the splits (Fowlkes and Mallows, 1983).

The comparison between the consensus transmission trees constructed from the simulated sequence data and the recorded infectious contacts further highlight some of the limitations of using a phylogenetic framework to infer transmission without careful

consideration of either epidemiological biases. These include the sampling strategy and features of pathogen evolution and the transmission processes, such as incomplete lineage sorting, that are known to influence phylogenetic reconstruction (de Maio *et al.* 2016). For example, in this study the inferred transmission trees were constructed using the sequence data from all infected sheds on the last day of the simulation; representing all infected hosts but only at a single time point during each outbreak. This is a major limitation as no inferences can be made for non-sampled hosts that would have contributed to the spread of infection earlier in the infection process. Although many studies have shown that low sampling coverage or differences in subsampling strategies do not contribute to substantial biases in other estimates such as the epidemic starting date (Ratmann *et al.* 2017; Hidano and Gate, 2019), it is important to try and ensure samples are taken from across the tree. Nevertheless, given the novelty of this model, it would be important for future studies to complete a sensitivity analysis in order to further describe these phenomena; for example, looking at the effect of transmission tree reconstruction under different sampling strategies.

A further sensitivity analysis could also be performed to help scale model parameters without having to rely on existing data. This may also have a large effect on the ability to infer who-infected-whom particularly with regards to the transmission rate which, if reduced, would help prevent the homogeneous cycle of infection that is seen in the current model; whereby each farm is repeatedly re-infected; making it difficult to identify a single chain of transmission - particularly given the limited number of SNPs accumulated throughout an outbreak. The use of a different model pathogen may also be helpful in refining both the model structure and the analytical methods; *C. jejuni* is a relatively slow evolving pathogen, resulting in very little genetic diversity over the length of the simulation, making it more difficult to correctly infer transmission dynamics (Frost *et al.* 2015; Rife *et al.* 2017). Many previous studies have highlighted the same limitation for other slowly evolving pathogens; for example, in a study by Campbell and colleagues

(2018) the comparison of two different outbreak reconstruction tools, *outbreaker* and *phybreak*, highlighted how pathogens with a lower mean transmission divergence provide little to no information about individual transmission events, and there is a need to expand phylogenetic tools to integrate epidemiological data that may help improve inferences (Campbell *et al.* 2018).

This use of *phybreak* has previously been compared to other outbreak reconstruction tools that can be implemented using the R packages *outbreaker* (Jombart *et al.* 2014) and *TransPhylo* (Didelot *et al.* 2017). Comparisons show that *phybreak* is able to correctly identify a greater number of transmissions including in outbreaks that are relatively small or those with different generation and sampling interval distributions (Klinkenberg *et al.* 2017). However, further comparison against different computational methods for the reconstruction of transmission trees, suggests that *phybreak* may not always be the most appropriate model. For example, in a study by Firestone and colleagues (2019) *phybreak* was shown to have the highest accuracy in comparison to five other models but only when considering those sources with consensus support which included just 6.5% of the inferred sources (Firestone *et al.* 2019). This supports the use of *phybreak* in relatively small outbreaks, however other models have been shown to have a similar predictive accuracy for larger datasets, such as those produced by the current model simulations. These include the SCOTTI model which models each host as a distinct population and transmissions between hosts as migration events (de Maio *et al.* 2016), and the Lau model that has been specifically developed to accommodate individual-level spatio-temporal data (Lau *et al.* 2017). Therefore, to improve the transmission tree reconstruction it may also be important to test different models, such as the SCOTTI or Lau model, as well as changing the sampling strategy.

When considering these results, it is also important to consider both the appropriateness of the analytical methods and the model assumptions that will have influenced the simulated sequences before phylogenetic reconstruction. For example, a major

simplification is that only a single infection can occur within each shed. In reality there is often multiple strains present in a single production area and even two competing variants of the same strain (Newell and Fearnley, 2003) although infection of a single chicken with more than one strain of *Campylobacter* is a rare observation (Sahin *et al.* 2002) and many farms appear to be dominated by a single strain (Ring *et al.* 2005). Nevertheless, the presence of multiple infections is more likely to hinder phylogenetic reconstruction; particularly if sampling is limited and strains that are present end up not being sampled. The competition between strains will also affect the transmission and survival of the pathogen; for example, in a study by Bull and colleagues (2005) longitudinal sampling in UK broiler flocks showed the replacement and extinction of different *Campylobacter* strains over time (Bull *et al.* 2005). Other simplifications made in the model will also have influenced the degree of similarity between the simulated sequences, including a disregard for evolutionary complexities such as recombination, within host evolution, changing mutations rates and hyper-variable regions. For example, recombination is known to play a fundamental role in the genetic diversity of *Campylobacter*; generating diversity at twice the rate of *de novo* mutation and facilitating gene flow between *C. jejuni* and other *Campylobacter* strains (Fearnhead *et al.* 2005; Wilson *et al.* 2008). Therefore, disregarding this process in the model could have a significant impact on phylogenetic reconstruction and transmission inferences.

Other model assumptions have little effect on the sequence simulation but instead impact the transmission dynamics. For example, seasonal variation in the prevalence of *C. jejuni* has been reported not only in chickens (Wallace *et al.* 1997), but also other farm animals (Stanley *et al.* 1998) which coincides with a marked increase in human campylobacteriosis reports. These increases in prevalence usually occur during the summer months (Nylen *et al.* 2002; Meldrum *et al.* 2004) and may be particularly evident in certain sequence types (Friedrich *et al.* 2016), which would be an important consideration to model. In addition to seasonality, other important demographic features

have been overlooked that are thought to be important risk factors for *Campylobacter*; including differences in biosecurity level, and farm type (*i.e.*, free-range versus barn and organic flocks versus conventionally farmed) (Van Overbeke *et al.* 2006). Of particular importance would also be the inclusion of thinning events, whereby birds are reared to the maximum stocking density permitted before a proportion is removed to lower the density. This can take place several times before all the birds are finally removed from the shed and the next production cycle begins, and it has been shown that the thinning process is a major risk factor for the introduction of *Campylobacter* (Goddard *et al.* 2013; Koolman *et al.* 2014).

Another potential risk factor that has not been included in the current model is carryover; that is, if the farm had been infected in its previous production cycle, it may be re-infected in the start of the next cycle due to unremoved environmental contamination (Newell and Fearnley, 2003; Bronowsk *et al.* 2014). Although, the importance of carryover is highly debated, with many genotyping studies suggesting it is a relatively infrequent event (Evan and Sayers, 2000; Shreeve *et al.* 2002), the exclusion of this potential transmission route as well as the other production features described above, may change the modelled transmission dynamics. It is also important to consider the impact of making inferences on the farm demographics that were included in the model, such as the length of a farm's production cycle and downtime period. In order for future models to be able to make accurate inferences about the spread of *C. jejuni*, or any pathogen, through the New Zealand commercial broiler industry, it would be important to collect and update more basic demographic information directly from producers; however, for the purpose of this study making these inferences should provide a sufficiently realistic framework to investigate the use of pathogen phylogenies in an epidemiological investigation.

6.6. Conclusion

Despite the many limiting assumptions both in the disease simulation model and phylogenetic analysis, this study has provided an opportunity to gain insight into how a network-based disease transmission model can be combined with a phylogenetic analysis and used to help determine who-infected-whom and the relative contribution of different transmission pathways in the New Zealand commercial poultry industry. Results show that the association between the genetic relatedness of pathogens isolates and different distance matrices may not be enough to accurately infer disease transmission dynamics. Nevertheless, model results emphasise the importance of combining both network and local contacts in disease models.

6.7. Acknowledgements

The authors are grateful for the support provided by the staff of the Poultry Industry Association and the Egg Producers' Federation of New Zealand. This study was also supported by the Ministry for Primary Industries, Massey University School of Veterinary Science and the Biological Heritage National Science Challenge. In addition, the authors thank Arata Hidano and Jun-Hee Han for their supports and guidance in modelling using the C programming language.

6.8. References

- Acheson D, Allos BM. *Campylobacter jejuni* infections: update on emerging issues and trends. *Clin. Infect. Dis.* 32, 1201–1206, 2001
- Bloomfield S, Vaughan T, Benschop J, Marshall J, Hayman D, Biggs P, Carter P, French N. Investigation of the validity of two Bayesian ancestral state reconstruction models for estimating *Salmonella* transmission during outbreaks. *PLoS One* 14, e0214169, 2019
- Bronowski C, James CE, Winstanley C. Role of environmental survival in transmission of *Campylobacter jejuni*. *FEMS Microbiol. Lett.* 356, 8–19, 2014
- Bull SA, Allen VM, Domingue G, Jorgensen F, Frost JA, Ure R, Whyte R, Tinker D, Corry JEL, Gillard-King J, Humphrey TJ. Sources of *Campylobacter* spp. colonizing housed broiler flocks during rearing. *Appl. Environ. Microbiol.* 72, 645–652, 2006
- Campbell F, Strang C, Ferguson N, Cori A, Jombart T. When are pathogen genome sequences informative of transmission events? *PLoS Pathog.* 14, e1006885, 2018
- de Cao E, Zagheni E, Manfredi P, Melegaro A. The relative importance of frequency of contacts and duration of exposure for the spread of directly transmitted infections. *Biostatistics* 15, 470–483, 2014
- Cawthraw SA, Newell DG. Investigation of the presence and protective effects of maternal antibodies against *Campylobacter jejuni* in chickens. *Avian Dis.* 54, 86–93, 2010
- Clarke K, Gorley R. *PRIMER v7: User Manual/Tutorial*. PRIMER-E, Plymouth, UK, 2015
- Conlan AJK, Coward C, Grant AJ, Maskell DJ, Gog JR. *Campylobacter jejuni* colonization and transmission in broiler chickens: a modelling perspective. *J. R. Soc. Interface* 4, 819–829, 2007
- Cooper KK, Songer JG, Uzal FA. Diagnosing clostridial enteric disease in poultry. *J. Vet. Diagnost. Investigat.* 25, 314–327, 2013
- Cottam EM, Thebaud G, Wadsworth J, Gloster J, Mansley L, Paton DJ, King DP, Haydon DT. Integrating genetic and epidemiological data to determine transmission pathways of foot-and-mouth disease virus. *Proc. Biol. Sci.* 275, 887–95, 2008

- Cousins M, Sargeant JM, Fisman D, Greer AL.** Modelling the transmission dynamics of *Campylobacter* in Ontario, Canada, assuming house flies, *Musca domestica*, are a mechanical vector of disease transmission. *R. Soc. Open Sci.* 6, e181394, 2019
- Craft ME.** Infectious disease transmission and contact networks in wildlife and livestock. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, e20140107, 2015
- Craven SE, Stern NJ, Line E, Bailey JS, Cox NA, Fedorka-Cray P.** Determination of the incidence of *Salmonella* spp., *Campylobacter jejuni*, and *Clostridium perfringens* in wild birds near broiler chicken houses by sampling intestinal droppings. *Avian Dis.* 44, 715–720, 2000
- Csardi G, Nepusz T.** The igraph software package for complex network research. *InterJournal, Complex Systems*, 2006. Available at: <http://igraph.org>
- Didelot X, Fraser C, Gardy J, Colijn C.** Genomic infectious disease epidemiology in partially sampled and ongoing outbreaks. *Mol. Biol. Evol.* 34, 997–1007, 2017
- Ekdahl K, Normann B, Andersson Y.** Could flies explain the elusive epidemiology of campylobacteriosis? *BMC Infect. Dis.* 5, 11–18, 2005
- Sawah AA, Dahshan HM, Nahass S, Mawgoud AI.** Pathogenicity of *Escherichia coli* O157 in commercial broiler chickens, *J. Basic Appl. Sci.* 7, 620–625, 2018
- Evans SJ, Sayers AR.** A longitudinal study of *campylobacter* infection of broiler flocks in Great Britain. *Prev. Vet. Med.* 46, 209–223, 2000
- Fearnhead P, Smith NGC, Barrigas M, Fox A, French N.** Analysis of recombination in *Campylobacter jejuni* from MLST population data. *J. Mol. Evol.* 61, 333–340, 2005
- Firestone SM, Hayama Y, Bradhurst R, Yamamoto T, Tsutsui T, Stevenson MA.** Reconstructing foot-and-mouth disease outbreaks: a methods comparison of transmission network models. *Scientific Reports* 9, 1–12, 2019
- Firestone SM, Ward MP, Christley RM, Dhand NK.** The importance of location in contact networks: describing early epidemic spread using spatial social network analysis. *Prev. Vet. Med.* 102, 185–195, 2011
- Fowlkes EB, Mallows CL.** A method for comparing two hierarchical clusterings. *J. Am. Stat. Assoc.* 78, 553–569, 1983

- French NP, Zhang J, Carter GP, Midwinter AC, Biggs PJ, Dyet K, Gilpin BJ, Ingle DJ, Mulqueen K, Rogers LE, Wilkinson DA, Greening SS, Mullner P, Williamson DA. Rapid and concurrent emergence of fluoroquinolone and tetracycline-resistant *Campylobacter jejuni* in humans and poultry in New Zealand. *Emerg. Infect. Dis.* 25, 2226–2235, 2019
- Friedrich A, Marshall JC, Biggs PJ, Midwinter AC, French NP. Seasonality of *Campylobacter jejuni* isolates associated with human campylobacteriosis in the Manawatu region, New Zealand. *Epidemiol. Infect.* 144, 820–828, 2016
- Frost SDW, Pybus OG, Gog JR, Viboud C, Bonhoeffer S, Bedford T. Eight challenges in phylodynamic inference. *Epidemics* 10, 88–92, 2015
- Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. *Software-Practice and Experience* 21, 1129–1164, 1991
- Galili T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 31, 3718–3720, 2015
- Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodtkin E, Rempel S, Moore R, Zhao Y, Holt R, Varhol R, Birol I, Lem M, Sharma MK, Elwood K, Jones SJ, Brinkman FS, Brunham RC, Tang P. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N. Engl. J. Med.* 364, 730–9, 2011
- Gates MC, Woolhouse MEJ, Gunn GJ, Humphry RW. Relative associations of cattle movements, local spread, and biosecurity with bovine viral diarrhoea virus (BVDV) seropositivity in beef and dairy herds. *Prev. Vet. Med.* 112, 285–295, 2013
- van Gerwe T, Mifflin JK, Templeton JM, Bouma A, Wagenaar JA, Jacobs-Reitsma WF, Stegeman A, Klinkenberg D. Quantifying transmission of *Campylobacter jejuni* in commercial broiler flocks. *Appl. Environ. Microbiol.* 75, 625–628, 2009
- Goddard AD, Arnold ME, Allen VM, Snary EL. Estimating the time at which commercial broiler flocks in Great Britain become infected with *Campylobacter*: a Bayesian approach. *Epidemiol. Infect.* 142, 1884–1892, 2014
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075, 2013

- Hall MD, Woolhouse MEJ, Rambaut A.** The effects of sampling strategy on the quality of reconstruction of viral population dynamics using Bayesian skyline family coalescent methods: A simulation study. *Virus Evol.* 2, e003, 2016
- Hayama Y, Firestone SM, Stevenson MA, Yamamoto T, Nishi T, Shimizu Y, Tsutsui T.** Reconstructing a transmission network and identifying risk factors of secondary transmissions in the 2010 foot-and-mouth disease outbreak in Japan. *Trans. Emerg. Dis.* 66, 2074–2086, 2019
- Hidano A, Gates MC.** Assessing biases in phylodynamic inferences in the presence of super-spreaders. *Vet. Res.* 50, 74–92, 2019
- Hijmans RJ.** *geosphere: Spherical Trigonometry.* R package version 1.5-10. <https://CRAN.R-project.org/package=geosphere>. 2019
- Humphrey T, O’Brien S, Madsen M.** *Campylobacters* as zoonotic pathogens: A food production perspective. *Int. J. Food Microbiol.* 117, 237–257, 2007
- Hutchison ML, Walters LD, Moore A, Crookes KM, Avery SM.** Effect of Length of Time before Incorporation on Survival of Pathogenic Bacteria Present in Livestock Wastes Applied to Agricultural Soil. *Appl. Environ. Microbiol.* 70, 5111–5118, 2004
- Jacobs-Reitsma W, Lyhs U, Wagenaar J.** *Campylobacter in the Food Supply.* In Nachamkin I, Szymanski C, Blaser M (ed), *Campylobacter*, Third Edition. p627-644. ASM Press, Washington, DC, 2008
- Jombart T, Cori A, Didelot X, Cauchemez S, Fraser C, Ferguson N.** Bayesian reconstruction of disease outbreaks by combining epidemiologic and genomic data. *PLoS Comp. Biol.* 10, e1003457, 2014
- Jukes TH, Cantor CR.** *Evolution of protein molecules.* in mammalian protein metabolism. p21-132. Academic Press, New York, 1969
- Kaakoush NO, Castaño-Rodríguez N, Mitchell HM, Man SM.** Global epidemiology of *Campylobacter* infection. *Clin. Microbiol. Rev.* 28, 687–720, 2015
- Klinkenberg D, Backer JA, Didelot X, Colijn C, Wallinga J.** Simultaneous inference of phylogenetic and transmission trees in infectious disease outbreaks. *PLoS Comput. Biol.* 13, e1005495, 2017

- Koolman L, Whyte P, Bolton DJ.** An investigation of broiler caecal *Campylobacter* counts at first and second thinning. *J. Appl. Microbiol.* 117, 876–881, 2014
- Kumar S, Stecher G, Tamura K.** MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874, 2016
- Lau MSY, Gibson GJ, Adrakey H, McClelland A, Riley S, Zelner J, Streftaris G, Funk S, Metcalf J, Dalziel BD, Grenfell BT.** A mechanistic spatio-temporal framework for modelling individual-to-individual transmission-With an application to the 2014-2015 West Africa Ebola outbreak. *PLoS Comput. Biol.* 13, e1005798, 2017
- Lessler J, Salje H, Grabowski MK, Cummings DAT.** Measuring spatial dependence for infectious disease epidemiology. *PLoS ONE* 11, e0155249, 2016
- Leventhal GE, Kouyos R, Stadler T, Wyl V von, Yerly S, Böni J, Cellerai C, Klimkait T, Günthard HF, Bonhoeffer S.** Inferring epidemic contact structure from phylogenetic trees. *PLoS Comput. Biol.* 8, e1002413, 2012
- Lin J.** Novel approaches for *Campylobacter* control in poultry. *Foodborne Pathog. Dis.* 6, 755–765, 2009
- de Maio ND, Wu C-H, Wilson DJ.** SCOTTI: efficient reconstruction of transmission within outbreaks with the structured coalescent. *PLoS Comput. Biol.* 12, e1005130, 2016
- Mantel N.** The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27, 209–220, 1967
- Meldrum RJ, Griffiths JK, Smith RMM, Evans MR.** The seasonality of human campylobacter infection and *Campylobacter* isolates from fresh, retail chicken in Wales. *Epidemiol. Infect.* 133, 49–52, 2005
- Metzig C, Ratmann O, Bezemer D, Colijn C.** Phylogenies from dynamic networks. *PLoS Comput. Biol.* 15, e1006761, 2019
- Mills A, Phillips CA.** *Campylobacter jejuni* and the human food chain: a possible source. *Nutrit. Food Sci.* 33, 197–202, 2003
- Muellner P, Kells N, Campbell D.** *Risk profile: the emergence of Campylobacter jejuni ST 6964 in poultry in New Zealand and its associated antimicrobial resistance.* MPI Technical Paper No: 2016/16, Wellington, New Zealand, 2016

- Neves MI, Malkawi I, Walker M, Alaboudi A, Abu-Basha E, Blake DP, Guitian J, Crotta M. The transmission dynamics of *Campylobacter jejuni* among broilers in semi-commercial farms in Jordan. *Epidemiol. Infect.* 147, e134, 2019
- Newell DG, Elvers KT, Dopfer D, Hansson I, Jones P, James S, Gittins J, Stern NJ, Davies R, Connerton I, Pearson D, Salvat G, Allen VM. Biosecurity-based interventions and strategies to reduce *Campylobacter* spp. on poultry farms. *Appl. Environ. Microbiol.* 77, 8605–8614, 2011
- Newell DG, Fearnley C. Sources of *Campylobacter* colonization in broiler chickens. *Appl. Environ. Microbiol.* 69, 4343–4351, 2003
- Nichols GL. Fly Transmission of *Campylobacter*. *Emerg. Infect. Dis.* 11, 361–364, 2005
- Nylen G, Dunstan F, Palmer SR, Andersson Y, Bager F, Cowden J, Feierl G, Galloway Y, Kapperud G, Megraud F, Molbak K, Petersen LR, Ruutu P. The seasonal distribution of *Campylobacter* infection in nine European countries and New Zealand. *Epidemiol. Infect.* 128, 383–390, 2002
- Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* 2, e000056, 2016
- Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528, 2019
- Qi L, Beaunée G, Arnoux S, Dutta BL, Joly A, Vergu E, Ezanno P. Neighbourhood contacts and trade movements drive the regional spread of bovine viral diarrhoea virus (BVDV). *Vet. Res.* 50, 30-42, 2019
- R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2018. Available at: <https://www.R-project.org/>
- Ratmann O, Hodcroft EB, Pickles M, Cori A, Hall M, Lycett S, Colijn C, Dearlove B, Didelot X, Frost S, Hossain ASMM, Joy JB, Kendall M, Kühnert D, Leventhal GE, Liang R, Plazzotta G, Poon AFY, Rasmussen DA, Stadler T, Volz E, Weis C, Leigh Brown AJ, Fraser C. Phylogenetic tools for generalized HIV-1 epidemics: findings from the PANGEA-HIV methods comparison. *Mol. Biol. Evol.* 34, 185–203, 2017

- Ridley A, Morris V, Gittins J, Cawthraw S, Harris J, Edge S, Allen V. Potential sources of *Campylobacter* infection on chicken farms: contamination and control of broiler-harvesting equipment, vehicles and personnel. *J. Appl. Microbiol.* 111, 233–244, 2011
- Rife BD, Mavian C, Chen X, Ciccozzi M, Salemi M, Min J, Prosperi MC. Phylodynamic applications in 21st century global infectious disease research. *Global Health Res. and Policy* 2, 13–25, 2017
- Ring M, Zychowska MA, Stephan R. Dynamics of *Campylobacter* spp. spread investigated in 14 broiler flocks in Switzerland. *Avian Dis.* 49, 390–396, 2005
- Robinson K, Fyson N, Cohen T, Fraser C, Colijn C. How the dynamics and structure of sexual contact networks shape pathogen phylogenies. *PLoS Comput. Biol.* 9, e1003105, 2013
- Rossi G, Leo GAD, Pongolini S, Natalini S, Zarengchi L, Ricchi M, Bolzoni L. The potential role of direct and indirect contacts on infection spread in dairy farm networks. *PLoS Comput. Biol.* 13, e1005301, 2017
- Sahin O, Morishita TY, Zhang Q. *Campylobacter* colonization in poultry: sources of infection and modes of transmission. *Anim. Health Res. Rev.* 3, 95–105, 2002
- Shreeve JE, Toszeghy M, Ridley A, Newell DG. The carry-over of *Campylobacter* isolates between sequential poultry flocks. *Avian Dis.* 46, 378–385, 2002
- Slader J, Domingue G, Jørgensen F, McAlpine K, Owen RJ, Bolton FJ, Humphrey TJ. Impact of transport crate reuse and of catching and processing on *Campylobacter* and *Salmonella* contamination of broiler chickens. *Appl. Environ. Microbiol.* 68, 713–719, 2002
- Sokal RR, Rohlf FJ. The comparison of dendrograms by objective methods. *Int. Assoc. Plant Taxon.* 11, 33–40, 1962
- Spada E, Saggiocca L, Sourdis J, Garbuglia AR, Poggi V, De Fusco C, Mele A. Use of the minimum spanning tree model for molecular epidemiological investigation of a nosocomial outbreak of hepatitis C virus infection. *J. Clin. Microbiol.* 42, 4230–6, 2004
- Stanley KN, Wallace JS, Currie JE, Diggle PJ, Jones K. Seasonal variation of thermophilic *Campylobacters* in lambs at slaughter. *J. Appl. Microbiol.* 84, 1111–1116, 1998a

- Stanley KN, Wallace JS, Currie JE, Diggle PJ, Jones K.** The seasonal variation of thermophilic *Campylobacters* in beef cattle, dairy cattle and calves. *J. Appl. Microbiol.* 85, 472–480, 1998b
- Stern NJ, Cox NA, Musgrove MT, Park CM.** Incidence and levels of *Campylobacter* in broilers after exposure to an inoculated seeder bird. *J. Appl. Poultry Res.* 10, 315–318, 2001
- Tabler GT, Berry IL, Mendenhall AM.** Mortality patterns associated with commercial broiler production. *Avian Advice* 6, 1–5, 2004
- USDA APHIS Veterinary Services.** Highly Pathogenic Avian Influenza (HPAI) response ready reference guide-overview of zones. USDA APHIS Veterinary Services, Riverdale, Maryland, United States, 2017. Available at: www.aphis.usda.gov/animal_health/emergency_management/downloads/hpai/hpai_zones.pdf. Accessed on 20 November 2019.
- VanderWaal K, Enns EA, Picasso C, Alvarez J, Perez A, Fernandez F, Gil A, Craft M, Wells S.** Optimal surveillance strategies for bovine tuberculosis in a low-prevalence country. *Sci. Rep.* 7, e4140, 2017
- van Gerwe TJWM, Bouma A, Jacobs-Reitsma WF, van den Broek J, Klinkenberg D, Stegeman JA, Heesterbeek JAP.** Quantifying Transmission of *Campylobacter* spp. among Broilers. *Appl. Environ. Microbiol.* 71, 5765–5770, 2005
- van Overbeke I, Duchateau L, De Zutter L, Albers G, Ducatelle R.** A comparison survey of organic and conventional broiler chickens for infectious agents affecting health and food safety. *Avian Dis.* 50, 196–200, 2006
- Vasylyeva TI, Friedman SR, Paraskevis D, Magiorkinis G.** Integrating molecular epidemiology and social network analysis to study infectious diseases: Towards a socio-molecular era for public health. *Infect. Genetics and Evolut.* 46, 248–255, 2016
- Wagenaar JA, Jacobs-Reitsma W, Hofshagen M, Newell D.** Poultry colonization with *Campylobacter* and its control at the primary production level. In Nachamkin I, Szymanski C, Blaser M (ed), *Campylobacter*, Third Edition. p667-685. ASM Press, Washington, DC, 2008
- Wallace JS, Stanley KN, Currie JE, Diggle PJ, Jones K.** Seasonality of thermophilic *Campylobacter* populations in chickens. *J. Appl. Microbiol.* 82, 219–224, 1997

- van der Wielen PWJJ van der, Biesterveld S, Notermans S, Hofstra H, Urlings BAP, Knapen F van.** Role of volatile fatty acids in development of the cecal microflora in broiler chickens during growth. *Appl. Environ. Microbiol.* 66, 2536–2540, 2000
- Wilson DJ, Gabriel E, Leatherbarrow AJH, Cheesbrough J, Gee S, Bolton E, Fox A, Hart CA, Diggle PJ, Fearnhead P.** Rapid evolution and the importance of recombination to the gastroenteric pathogen *Campylobacter jejuni*. *Mol. Biol. Evol.* 26, 385–397, 2009
- Yassin H, Velthuis AGJ, Boerjan M, van Riel J.** Field study on broilers' first-week mortality. *Poultry Sci.* 88, 798–804, 2009
- Yin Q, Shi T, Dong C, Yan Z.** The impact of contact patterns on epidemic dynamics. *PLoS ONE* 12, e0173411, 2017

**Genetic relatedness of *Staphylococcus aureus*
isolates obtained from New Zealand dairy cattle**

Sabrina S. Greening, Ji Zhang, Anne C. Midwinter, David A. Wilkinson, Scott
McDougall, M. Carolyn Gates, and Nigel P. French

7.1. Abstract

AIMS: *Staphylococcus aureus* is one of the leading causes of bovine mastitis worldwide and is a common indication for use of antimicrobials on dairy farms. However, the use of antimicrobial agents on dairy farms, as well as in other food animal production systems, is a major concern for its role in the emergence of antimicrobial resistance (AMR) in *S. aureus*. The aims of this study were to (i) describe the genetic population structure of mastitis-causing *S. aureus* in a sample of New Zealand dairy herds, (ii) identify genes associated with virulence and AMR, (iii) investigate the association between on-farm antimicrobial usage and the AMR profiles of *S. aureus*, (iv) examine the potential contribution of live animal movements to the spread of *S. aureus* between the study herds and lastly, (v) to determine the relationship between the bovine-derived *S. aureus* isolates collected in this study and isolates collected previously from humans, domesticated pets and livestock both within New Zealand and internationally.

METHODS: Whole-genome sequencing was performed on a total of 57 *S. aureus* isolates derived from cows with either clinical or subclinical mastitis across 17 New Zealand dairy herds located in the Waikato region of the North Island. The genetic relatedness between the isolates was examined using the core single nucleotide polymorphism alignment to a construct phylogenetic tree whilst AMR and virulence genes were identified *in silico*. The association between on-farm antimicrobial usage and the presence of resistance genes was investigated. In addition, a permutational multivariate analysis of variance analysis model was used to examine the relative importance of live animal movements and the spatial proximity of farms as determinants of the pairwise genetic relatedness of *S. aureus*. Lastly, two maximum-likelihood phylogenetic trees were constructed to investigate the relationship between the bovine-derived *S. aureus* isolates collected in this study and (i) 59 *S. aureus* isolates collected previously from domesticated pets and humans in New Zealand and (ii) 103 *S. aureus* Sequence Type (ST)-1 isolates collected previously from across 19 different countries.

RESULTS: This study identified eight *S. aureus* Sequence Types (STs) in the 57 *S. aureus* isolates, of which 61.4% (35/57) belonged to ST-1. Altogether 14 genes associated with AMR and 76 genes associated with virulence factors were identified with little genetic diversity between isolates belonging to the same ST. Overall, no association ($p < 0.05$) was found between gene presence-absence and the use of four different dry cow therapy treatments. However, the genetic relatedness of isolates was found to be associated with both the trading community ($p = 0.0001$) and township ($p = 0.0004$). In the global comparison of ST-1 isolates, the New Zealand bovine isolates formed a monophyletic group with the exception of one closely related bovine-derived isolate originating from Australia; whilst other New Zealand isolates collected from both domesticated pets and humans were interspersed throughout the phylogeny; clustering with other international *S. aureus* isolates.

CONCLUSIONS: This study characterised bovine-derived *S. aureus* isolates from bovine mastitis cases in the Waikato region of New Zealand. Overall, the majority of isolates were related to historically human-derived STs, whilst STs thought to be dominant worldwide in bovine milk were relatively rare in our sample. No association was found between the AMR gene profiles and either on-farm antimicrobial usage or antimicrobial sensitivity patterns and, although results provide evidence that the movement of live animals may be an important risk factor for the regional spread of *S. aureus*, it is clear that using cattle-tracing data alone may not be enough to fully capture the transmission dynamics of *S. aureus* between farms due to other contacts that may be contributing to the spread of the pathogen.

KEYWORDS: *Staphylococcus aureus*, Network community, PERMANOVA, Whole-genome sequencing, Comparative genomics, Antimicrobial resistance

7.2. Introduction

Bovine mastitis continues to be one of the most economically important diseases affecting the dairy cattle industry worldwide despite intensive research and the increasing uptake of various on-farm control strategies (Seegers *et al.* 2003; Petrovski *et al.* 2006). With over 137 different organisms including bacteria such as mycoplasma, yeasts and algae having been previously described as causative agents (Watts, 1988), one of the major challenges in controlling bovine mastitis is correctly identifying the pathogen responsible for causing disease (Bogni *et al.* 2011). The relative importance of these different pathogens is largely country-dependent however, more than 90% of all new intra-mammary infections are caused by only a small number of pathogenic bacteria which are typically classified into two groups; contagious, comprising of *Streptococcus dysgalactiae*, *Streptococcus agalactiae* and *Staphylococcus aureus*, or environmental, comprising of *Escherichia coli* and *Streptococcus uberis* (Eberhart, 1986; Bradley, 2002). These classifications are not always clear cut but are based on the source of infection, mode of transmission and tendency to cause persistent or transient intra-mammary infection.

In the New Zealand dairy industry, the epidemiology of the predominant mastitis-causing pathogens varies greatly in comparison with many countries in the northern hemisphere, where there is a greater reliance on indoor housing systems and year-round calving practices. However, within the last five decades New Zealand has seen a notable change in the aetiology and epidemiology of mastitis in dairy cows (McDougall, 2002; Petrovski *et al.* 2009; Heffernan *et al.* 2015) with contagious mastitis pathogens, such as *S. agalactiae*, decreasing in prevalence in comparison to environmental pathogens, such as *S. uberis*, which are now on the rise (McDougall, 2002; McDougall *et al.* 2007; Petrovski *et al.* 2011). This change followed the successful uptake of a mastitis control strategy entitled the Seasonal Approach to Managing Mastitis (SAMM plan), which operated in New Zealand between 1993 and 2010, and its subsequent development into the SmartSAMM programme that encouraged antibiotic dry-cow therapy for the treatment

of intra-mammary infection at the end of lactation (Lacy-Hulbert *et al.* 2011). However, current recommendations are moving away from the treatment of all cows with antimicrobials at the end of lactation (*i.e.*, “blanket dry cow therapy”) towards more targeted dry cow therapy and the use of bismuth subnitrate-based internal teat sealants, with a goal to eliminate the use of antibiotics for the maintenance of animal health and wellness across New Zealand by 2030 (Hillerton and Allison, 2015).

Despite the overall decline in contagious mastitis-causing pathogens, *S. aureus* continues to cause significant economic losses in the dairy industry due largely to its role in sub-clinical and chronic disease resulting in its long-term persistence within many dairy herds (Rall *et al.* 2014; Bonsaglia *et al.* 2018; Rossi *et al.* 2019). In addition, many New Zealand *S. aureus* isolates collected from sub-clinical mastitis cases also show resistance to common antibiotic classes used in dry-cow therapy treatments (McDougall *et al.* 2014; Petrovski *et al.* 2015). This growing pattern of resistance raises concerns over the use of antimicrobials for the treatment and control of mastitis, due to not only the general concern of using antibiotics in food-producing livestock but also because there is increasing evidence for livestock-associated methicillin-resistant *S. aureus* (LA-MRSA) in humans cases (Mehndiratta and Bhalla, 2014; Cuny *et al.* 2015; Mohammed and Nigatu, 2015).

Since the first LA-MRSA was described in Belgium in 1972, isolated from a case of bovine mastitis (Devriese *et al.* 1972), there has been a series of studies looking at the emergence, evolution and dissemination of LA-MRSA. Many studies have focused on the most widely spread LA-MRSA clonal complex (CC) 398 that was first isolated from nasal swabs of breeding pigs (Williamson *et al.* 2013; Becker *et al.* 2015) but has subsequently been detected in numerous livestock species, including veal calves, poultry, dairy cattle and goats, as well as domesticated pets and human cases (Cuny *et al.* 2015; Gonçalves da Silva *et al.* 2017). The latter are usually limited to those in close contact with livestock species

(Wulf and Voss, 2008; Fluit 2012; Verkade *et al.* 2013). Due to this growing concern of LA-MRSA, an increasing number of studies are using molecular typing methods such as Pulsed-Field Gel Electrophoresis (PFGE) and Multilocus Sequence Typing (MLST) to look at the genetic relationships between *S. aureus* isolates. These techniques provided higher resolution genotypic insights into the transmission and evolutionary dynamics of *S. aureus* between animal and human populations; supporting evidence for both the emergence of historically livestock-associated *S. aureus* strains in humans (Spoor *et al.* 2013; Costa *et al.* 2015) and animal-adapted strains thought to be derived from humans within livestock populations (Lowder *et al.* 2009; Guinane *et al.* 2010; Sakwinska *et al.* 2011; Price *et al.* 2012). However, there is limited evidence for presence of MRSA in livestock in New Zealand (Grinberg *et al.* 2008), and a very low prevalence of LA-MRSA isolates in human derived isolates (Heffernan *et al.* 2015).

In addition to molecular typing, many studies have used contacts networks to determine which interactions within a population are most important for the introduction, spread and persistence of *S. aureus* strains (Ciccolini *et al.* 2012; Jarynowski and Liljeros, 2015; Obadia *et al.* 2015; Pei *et al.* 2018). Contact networks are constructed from a set of elements, often referred to as nodes, vertices or actors, which represents a unit of interest within the network such as a farm, with edges capturing contacts between any two nodes in the network, such as the movement of animals between two farms (Wasserman and Faust, 1994). Network models are a popular tool in infectious disease epidemiology; providing insight into the transmission of infectious diseases by helping to determine individuals in the network that pose the greatest risk for disease dissemination and informing disease control and surveillance activities (Christley *et al.* 2005; Bansal *et al.* 2007; Danon *et al.* 2011). More recently, an increasing number of network models have used molecular typing data to help estimate model parameters and look at both transmission and evolutionary dynamics. However, one limitation in building these models is the requirement to have both adequate network data and pathogen molecular

data. For this reason, examples integrating *S. aureus* molecular data with livestock contact networks is limited, although one study by Álvarez and colleagues (2011) is a good example of how these data sources can be integrated to study the transmission and evolution of *S. aureus* strains, and there are several studies that have looked at other pathogens in cattle networks (Biek *et al.* 2012; Broeckl *et al.* 2017; de Knecht *et al.* 2018). The Álvarez and colleagues (2011) study used a network-based model constructed from the reported movements of cattle, and molecular data on bovine *S. aureus*, to highlight the importance of cattle movements and other local contacts, such as farm visitors, for circulating different *S. aureus* strains between farms (Álvarez *et al.* 2011).

In New Zealand, despite there being a number of studies have looked at the genetic diversity, spatial distribution and antimicrobial susceptibility patterns of *S. aureus* strains in cattle (McDougall *et al.* 2014; Petrovski *et al.* 2015), few attempts have been made to either integrate contact networks with molecular data, to investigate the contribution of cattle trade on the spread of *S. aureus*, or to look at the genetic relationship between human and cattle *S. aureus* isolates to investigate the role of cross-zoonotic transmission on the distribution and persistence of strains in New Zealand; with common, clonal aetiology previously shown between farmers and cows in New Zealand (Grinberg *et al.* 2004). This study aims to describe the genetic population structure of bovine mastitis-causing *S. aureus* from a small sample of New Zealand dairy herds; providing a snapshot of the genetic basis for virulence and resistance currently within the population. By further investigating the genetic relationship between these bovine mastitis-causing *S. aureus* isolates and human isolates, collected both throughout New Zealand and internationally, this study also aims to determine the potential for zoonotic transmission. Lastly, this study aims to investigate the role of cattle trade on the genetic relationship of bovine mastitis-causing *S. aureus* in order to consider the contribution of live animal movements in the spread and evolution of this important pathogen.

7.3. Materials and Methods

7.3.1. Current study: sample selection, microbiology, and whole-genome sequencing

The *S. aureus* isolates used in the present study were obtained from milk samples that were collected during a study conducted between 22nd October 2015 and 27th January 2016 across dairy herds located in the Waikato region of New Zealand's North Island. Herds were selected on a convenience basis (*i.e.*, the herd owners were willing to be involved in the study). Cows with grossly evident changes to the milk and/or heat and swelling of the mammary gland (*i.e.*, clinical mastitis), or cows with an elevated somatic cell count (*i.e.*, >200,000 cells/mL) at production recording had milk samples collected following aseptic teat end preparation. Milk samples were submitted for routine microbiology following the procedures of the National Mastitis Council, US. For this current study, a subset of the *S. aureus* isolates were selected for whole-genome sequencing using stratified random sampling with three hierarchical groups created to ensure that there was (i) at least one isolate from each herd to explore the between-herd variation in *S. aureus* isolates, (ii) multiple isolates from different cows in the same herd to explore the within-herd variation in *S. aureus* isolates and, (iii) multiple isolates from different quarters on individual cows to explore the within-animal variation in *S. aureus* isolates. Before sampling the isolates, the study was judged to be low risk thorough peer evaluation and consequently was not formally reviewed by any of the University's Human Ethics Committees.

The selected *S. aureus* isolates were delivered on Dorset egg slopes to the ^mEpiLab, Massey University. All isolates were re-cultured on Columbia horse blood agar plates, and from a pure sub-culture a heavy inoculum was made in nutrient broth No. 2 (Oxoid, Hampshire, UK) containing 15% glycerol and an aliquot frozen at -80°C. From this, a sub-culture was made on Columbia horse blood agar and a single bacterial colony selected for DNA extraction using a QIAamp DNA Mini Kit (QIAGEN). DNA libraries were then prepared using the Nextera XT DNA preparation kit (Illumina, San Diego, USA) for

submission to New Zealand Genomics Limited (University of Otago, Dunedin, New Zealand), which performed 2 x 100 bp sequencing on the Illumina NextSeq 500 platform following the manufacturer's instructions.

7.3.2. Genomic analyses: current study

Raw reads from Illumina sequencing were evaluated, assembled, annotated, and analysed using the Nullarbor pipeline (v2.0.0) (<https://github.com/tseemann/nullarbor>). In short, the adapter sequences were trimmed with Trimmomatic (v0.38) (Bolger *et al.* 2014). Trimmed sequences subsequently underwent *de novo* assembly in SKESA (v2.3) (Souvorov *et al.* 2018), annotation in Prokka (v1.13) (Seemann, 2014) and scanned with PubMLST *S. aureus* typing scheme with MLST (<https://github.com/tseemann/mlst>). To generate an alignment of core genome single nucleotide polymorphisms (SNPs), sequence reads for each isolate were aligned to the reference genome MSSA476, a methicillin-susceptible *S. aureus* (MSSA) isolate assigned to ST-1 (GenBank Accession: NC_002953.3) within the Nullarbor pipeline using Snippy (v4.0) (<https://github.com/tseemann/snippy>). Recombinant regions were eliminated from the genome alignment and the remaining polymorphic sites were identified using the software tool Gubbins (v2.3.2) (Croucher *et al.* 2015) before removing indels and extracting SNP sites using the software tool SNP-sites (v2.4.0) (Page *et al.* 2016).

Identification of antimicrobial resistance (AMR) genes were also performed within the Nullarbor pipeline using the software tool ABRicate (v0.8) (<https://github.com/tseemann/abricate>) to screen contigs through ResFinder (v3.1) (Zankari, *et al.* 2012) and the Comprehensive Antibiotic Resistance Database (CARD) (Jia *et al.* 2017) whilst screening for virulence genes was undertaken via the Virulence Factors Database (Chen *et al.* 2005). Genome Profiler (GeP) (Zhang *et al.* 2015) was used to convert assembly data into whole-genome multilocus sequence typing (wgMLST) allelic profiles. The relationship between the isolates was then examined by using the core

SNP alignment to construct a maximum-likelihood phylogeny using the R package *ape* (Paradis and Schliep, 2018). Phylogenetic trees were displayed alongside the AMR and virulence gene profile of each isolate using the online tool Interactive Tree of Life (iTOL) (v4.5.3) (Letunic and Bork, 2016).

7.3.3. Antimicrobial sensitivity

The sensitivity of a subset of *S. aureus* isolates (50/57, 87.7%) was determined using a zone diffusion test following the procedures provided by the Clinical and Laboratory Standards Institute (CLSI). The antimicrobials assessed included penicillin (10 µg), novobiocin (5 µg), cefoxitin (30 µg), tetracycline (30 µg), ceftiofur (30 µg), and oxacillin (1 µg); with isolates being declared sensitive, intermediate or resistant, based on CLSI recommendations. The sensitivity of each isolate was also displayed alongside the phylogenetic tree using iTOL (v4.5.3) (Letunic and Bork, 2016) to examine the association between gene presence-absence and the sensitivity to each antimicrobial.

7.3.4. Comparative genomics: *S. aureus* within New Zealand

To determine the similarity of epidemiologically unrelated *S. aureus* isolates from across New Zealand, a multispecies comparative genomic analysis was performed with the addition of 59 MSSA isolates sampled from human clinical cases and colonised humans, dogs and cats across New Zealand. These isolates were collected as part of a previous study by Grinberg and colleagues (2017) investigating the genetic differentiation across colonisation and skin and soft tissue infections (SSTI) between people and household pets (Grinberg *et al.* 2017). Raw reads were accessed via the European Bioinformatics Institute online database (<https://www.ebi.ac.uk/ena>) using the accession number: PRJNA391123. The raw reads were processed using the Nullarbor pipeline described above followed by GeP analysis (Zhang *et al.* 2015) in order to obtain the wgMLST allelic profiles. A maximum-likelihood phylogeny was constructed from core SNPs to show the relationship between the 57 *S. aureus* isolates sampled from bovine milk for this study

and the 59 *S. aureus* isolates sampled from both humans and pets for the study by Grinberg and colleagues (2017). All figures depicting phylogenetic relationships and associated variables were created using the online tool Microreact (Argimón *et al.* 2016).

7.3.5. Comparative genomics: *S. aureus* ST-1 globally

To compare ST-1 genomes sequences in our study to the other previously submitted ST-1 genomes, all the available *S. aureus* genome assemblies ($n = 9,897$, accessed on 12.03.2019) were download from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>). The ST of the assemblies where scanned using the MLST pipeline (<https://github.com/tseemann/mlst>) and an additional multispecies comparative genomic analysis was performed looking at the genetic similarity between all the ST-1 *S. aureus* isolates including all the ST-1 isolates from the current study (the predominant ST), all the ST-1 whole-genome sequence assemblies identified from GenBank and, all the ST-1 isolates identified in the previous study by Grinberg and colleagues (2017). A list of the Genbank accession numbers of all the genome sequences used can be found in Appendix E (Table E1). Similarly, to both the regional and within-country genomic analyses, the raw reads were processed using the Nullarbor pipeline and the *wgMLST* analysis was performed with GeP (Zhang *et al.* 2015). A maximum-likelihood phylogeny was constructed from core SNPs showing the relationship between the 35 *S. aureus* ST-1 isolates from this study and 103 *S. aureus* ST-1 isolates found in GenBank. All figures depicting phylogenetic relationships were created using the online tool Microreact (Argimón *et al.* 2016).

Scoary (v 1.6.16), a software tool used for studying the association between gene presence-absence and known traits (Brynildsrud *et al.* 2016), was used to investigate the association between gene presence-absence and both sample location on a continental scale (*i.e.*, Asia, Africa, Europe, South America, North America, and Oceania), and for the sampled host (*i.e.*, bovine, human and household pets). These Scoary analyses were limited to only

ST-1 isolates in an attempt to control for the potential confounding effects of ST. In order to further control confounding by lineage it is recommended to use a pairwise comparison algorithm that accounts for contrasting pairs that share a common ancestor (Maddison, 2000; Brynildsrud *et al.* 2016). However, in these analyses, no causal inference *i.e.*, which genes cause membership in a group is being made, instead the aim of the analysis was to identify if any genes are overrepresented, therefore no pairwise comparison was made. Instead, Scoary results were filtered so that only genes that had an odds ratio greater than one, a specificity greater than 95%, a Benjamini-Hochberg corrected *p*-value (Benjamini and Hochberg, 1995) below 0.05, and were not annotated as either a “hypothetical protein” or a “putative protein” were considered significant. The remaining genes were subsequently used as search terms in the UniProt Knowledgebase (UniProtKB; <http://www.uniprot.org>) (The UniProt Consortium, 2011) to determine if genes had Gene Ontology (GO) terms (Gene Ontology Consortium, 2010) describing either biological processes or molecular functions that have been associated with the gene products (Dimmer *et al.* 2012).

7.3.6. Cattle trade network

To characterise the contact patterns between the 17 dairy farms included in our study, we obtained an extract of animal movement records from the Livestock Improvement Corporation (LIC) over the period from 1985 to 2011. LIC is a multinational farmer-owned co-operative that provides members with the computerised herd management software called MINDA, which can be used to record information on individual dairy cattle that have entered into lactation including basic production data as well as information on the purchases, sales, culls, and deaths of individual animals. Movement records for individual animals contain the movement date, the reason for the movement, and if a change of ownership occurred following the movement. Movement data was exported from Microsoft Excel into the R statistical software package (R Core Team, 2018) in order to construct a contact network to determine which of the 17 study farms

had been in contact with each other through the movement of at least one animal. The farms in our study were identified in the MINDA database using their unique LIC herd identification code. The network graph was constructed using the R package *igraph* (Csardi and Nepusz, 2006) with nodes representing all the farms that reported a movement in the MINDA database and directed edges between a pair of nodes representing the movement of a live animal. Edges were weighted to show if more than one animal movement had occurred. Common network metrics, described in Appendix E (Table E2), were calculated including measures to indicate network size (*i.e.*, the total number of nodes, the total number of edges and network diameter), centrality (*i.e.*, in- and out-degree and betweenness) and cohesion (*i.e.*, network density and clustering coefficient).

A fast-greedy community analysis was also completed based on the network community algorithm presented by Clauset and colleagues (2004) to identify groups of farms that have more internal links between them than external links to other communities within the network (Clauset *et al.* 2004). The community analysis was performed using the R package *igraph* (Csardi and Nepusz, 2006) on the complete network containing all the farms that reported a movement in the MINDA. Network community was then included as a variable in the permutational multivariate analysis of variance (PERMANOVA) model described below.

7.3.7. Antimicrobial usage

Antimicrobial usage (AMU) based on the daily dose per cow per lactation for each of the study herds was estimated from antimicrobial purchase data for the 2015-2016 lactation. This timeframe was chosen because this is the lactation from which the isolate samples were collected. The daily dose (DD) was calculated for intra-mammary treatments based on tubes per day; for example, if an intra-mammary treatment states two tubes per day on the label and the farm purchased a box of 20 tubes then it was recorded as 10 DDs. The

exception to this was dry cow therapy where four tubes (*i.e.*, one tube per quarter at drying off) is considered 1 DD. A similar method was used for injectables but instead the recommended volume (in millilitres) per cow per day was considered as the DD and this was divided by the total volume of products sold. For example, a treatment of 20 ml per cow per day would be considered the daily dose and if 100 ml of a product had been purchased then it was recorded as 5 DDs. An adjustment to the DD was made if the formulation was considered long acting. For example, some oxytetracycline treatments will have three days of activity in a single dose. Thus, a single treatment of 20 ml acting over three days would be equal to a DD of 7 ml. The DDs for all the antimicrobials purchased by a farm over the 12-month period was then summed and divided by the number of cows on-farm to give an average daily dose per cow per lactation.

To investigate the potential relationship between AMU and both herd size and the presence of resistance genes, scatterplots were created with AMU plotted against (i) the total number of animals per farm and (ii) the total number of resistance genes per farm; that is, the sum of all the unique resistance genes identified in each isolate sampled from that farm. A Spearman's rank correlation coefficient (SCC) could then be calculated using the R package *stats* (R-Core Team, 2018). A further Scoary analysis was then performed to investigate the association between the presence of resistance genes and the use of four common treatments used in dry cow therapy; BovacloxTM and Dryclox[®] (both containing 500 mg cloxacillin and 250 mg ampicillin), Cepravin[®] (containing 250 mg cephalonium), and Orbenin[®] (containing 500 mg of cloxacillin). For this analysis, the gene presence-absence matrix produced by Roary (v3.12.0) (Page *et al.* 2015) was used alongside a binary matrix indicating if an isolate had been sampled from a farm using any of the dry cow therapy treatments, such that "1" indicated the use of the treatment on-farm whilst "0" indicated that the treatment had not been used.

An additional Scoary analysis was also used to investigate the association between the presence of virulence genes and sample type (clinical versus sub-clinical). In this analysis, the same gene presence-absence matrix produced by Roary was used with the additional matrix indicating if isolates had been sampled from cows presenting with either clinical or sub-clinical mastitis. For both Scoary analyses, there was an interest in causal genes, therefore, unlike the previous Scoary analyses described above, a pairwise comparison was used in order to find the maximum number of phylogenetically non-intersecting pairs of isolates that contrast for both the gene and trait (Figure 7.1). Genes were considered significant if the entire range of pairwise comparison *p*-values were less than 0.05 with significant genes used as further search terms in the UniProtKB.

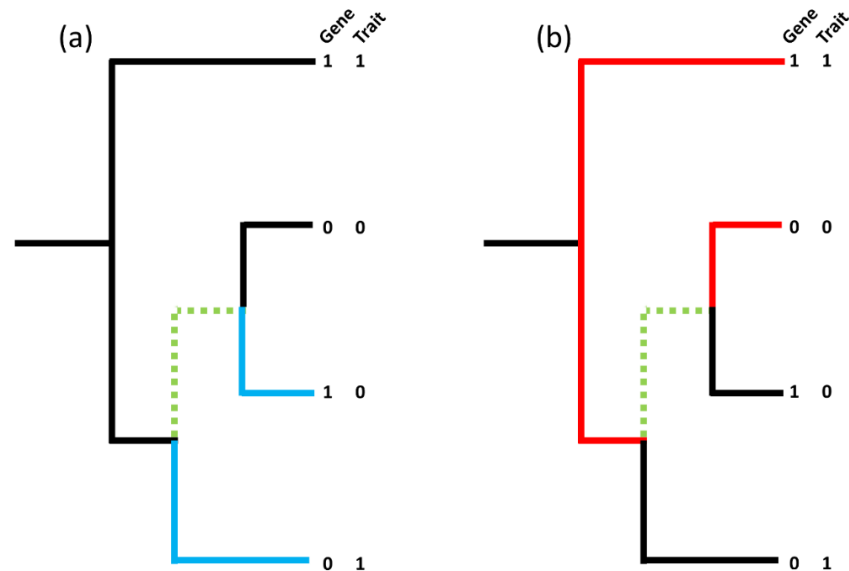


Figure 7.1. (a) shows one contrasting pair (blue branches; 1-0|0-1) whilst (b) shows another contrasting pair on the same tree (red branches; 1-1|0-0), however the maximum number of non-intersecting, contrasting pairs in the tree remains as one due to the common branches, highlighted in green, shared by both contrasting pairs. The “best” picking is the red pair with both the gene and the trait present whilst the “worst” picking is the blue pair. To handle confounding by lineage a pairwise comparisons algorithm can be used to identify the maximum number of non-intersecting, contrasting pairs in a tree and calculate the corresponding binomial test *p*-value taking into account the proportion of “best” and “worst” pickings over the set of non-intersecting, contrasting pairs.

7.3.8. Permutational multivariate analysis of variance

The core SNP alignment outputted from Gubbins (Croucher *et al.* 2015) was used to create a pairwise distance matrix in MEGA7 (v7.0.25) (Kumar *et al.* 2016) with values in the matrix corresponding to the uncorrected p -measure; that is, the proportion of nucleotide sites at which the two sequences being compared differ. The association between the genetic distance matrix and three independent variables; township, farm (as an independent variable and not nested in township), and network community, were evaluated using the PERMANOVA+ add-on package (Anderson *et al.* 2008) for PRIMER (v7.0). For this analysis, the farm's coordinates were used to determine which township they belonged to. The coordinates were first plotted using Google Maps (2017), and the township given in the address was recorded. Univariate PERMANOVA models were performed for each factor with p -values obtained using 9999 unrestricted permutations of raw data. Non-metric multidimensional scaling (MDS) ordination plots (Kruskal and Wish, 1978) were then constructed, also in PRIMER (v7.0), mapping isolates in a two-dimensional Euclidean space in a manner that preserves the genetic distance between each isolate. Plots were used to identify any outliers and further investigate the relationship between each variable and the genetic distance between each isolate by using different colours to identify isolates with a variable in common. To evaluate whether there was an association between two variables, as an indication for collinearity, a Chi-square test of independence was performed between each of the contingency tables using the R package *stats* (R-Core Team, 2018) before a final multivariate backward stepwise model was built with factors being removed if $p > 0.25$.

7.4. Results

7.4.1 Genomic analysis: current study

In total, 57 *S. aureus* isolates were selected for wgMLST analysis with isolates sampled from 51 cows with either clinical or sub-clinical mastitis on 17 farms (Table 7.1). The geographical distribution of farms was limited to the Waikato region of New Zealand's North Island (Figure 7.2) with the Euclidean distance between farms ranging from 0.7km

to 44.1km (mean = 19.9 km). Altogether, 33.3% (19/57) of the isolates were sampled from animals presenting with clinical mastitis compared to 66.7% (38/57) from sub-clinical animals. Overall, 51,796 polymorphic sites were identified amongst the 57 isolates containing 37,746 core SNPs. Eight unique STs were distinguished with the predominant sequence type, ST-1, being found across 64.7% (11/17) of the study farms and 61.5% (32/51) of the sampled animals (Table 7.1). The maximum-likelihood phylogeny is shown in Figure 7.2 however, the relatively large genetic distances between the major clades mask the fine-scale variation between isolates within clades, therefore a higher resolution phylogeny of the ST-1 cluster has been provided in Appendix E, Figure E1 with 1,063 polymorphic sites and 1,061 core SNPs identified between the 35 ST-1 isolates.

Table 7.1. Isolate-level (n = 57), farm-level (n = 17) and cow-level (n = 52) prevalence of eight *Staphylococcus aureus* whole-genome multi-locus sequence types (ST) isolated from the Waikato region of New Zealand North Island. The distribution of *S. aureus* sequence types among isolates show if the isolate was collected from a cow presenting with either clinical (n = 19) or sub-clinical (n = 38) mastitis.

	Multilocus sequence type (ST)							
	ST-1	ST-188	ST-5	ST-705	ST-1247	ST-97	ST-151	ST-425
Isolates								
(%)								
<i>Clinical</i>	12	0	0	1	0	2	1	0
	(63.2)	(0.0)	(0.0)	(5.3)	(0.0)	(10.5)	(5.3)	(0.0)
<i>Sub-clinical</i>	23	5	1	1	2	3	0	1
	(60.5)	(13.2)	(2.6)	(2.6)	(5.3)	(7.9)	(0.0)	(2.6)
<i>Total</i>	35	5	1	2	2	5	1	1
	(61.4)	(8.8)	(1.8)	(3.5)	(3.5)	(8.8)	(1.8)	(1.8)
Farms								
(%)								
	11	1	1	2	2	3	1	1
	(64.7)	(5.9)	(5.9)	(11.8)	(11.8)	(17.6)	(5.9)	(5.9)
Cows								
(%)								
	32	4	1	2	2	5	1	1
	(61.5)	(7.7)	(1.9)	(3.8)	(3.8)	(9.6)	(1.9)	(1.9)

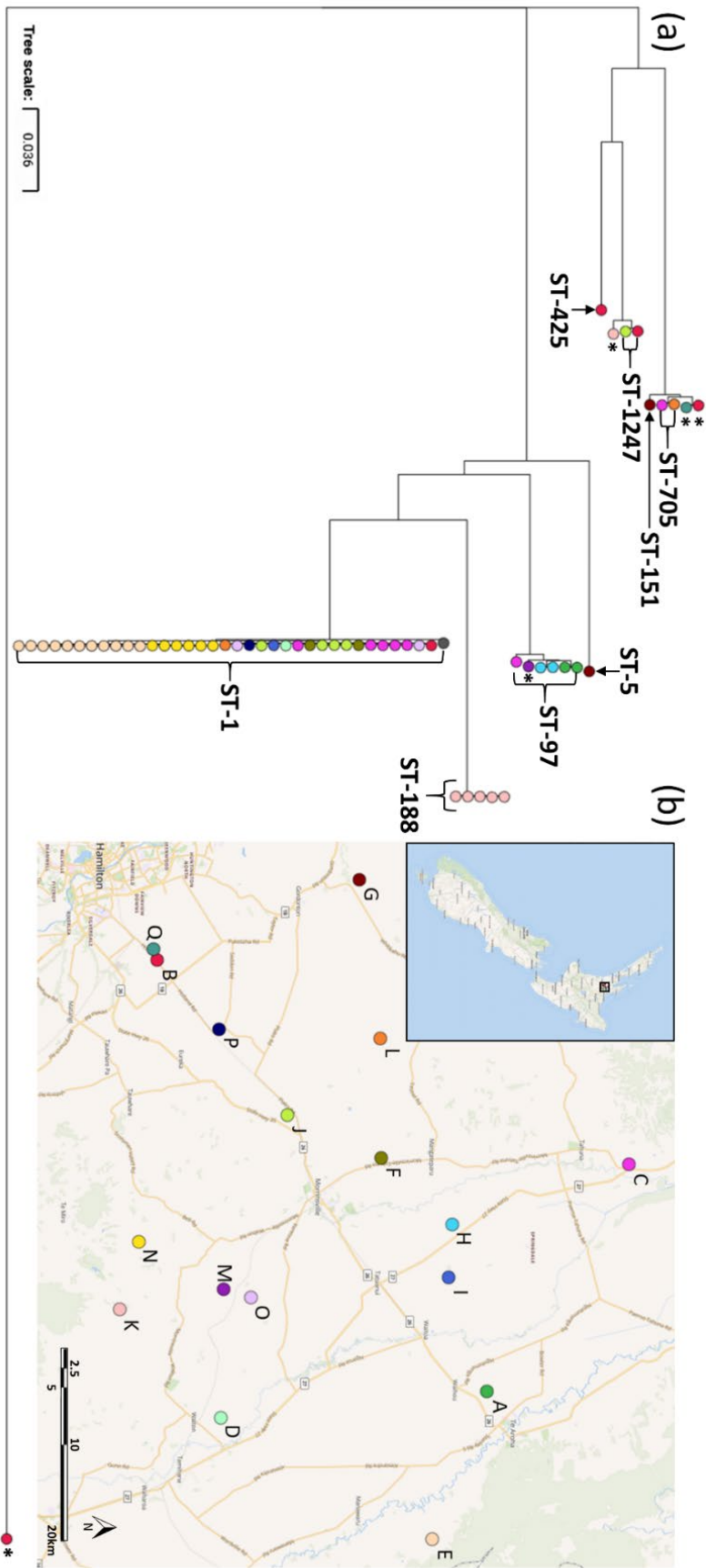


Figure 7.2. (a) Maximum-likelihood phylogeny generated from core single nucleotide polymorphisms across the 57 *S. aureus* isolates. The colour of the terminal tree nodes corresponds to the farm location shown in (b), and all isolates have been labelled with their sequence type identified by seven MLST genes. (b) Map showing the geographical distribution of selected dairy farms (n=17). The alphabetic labels (A-Q) have been used to identify the farms in the subsequent phylogenetic analysis. Figure created using the online tool Microreact (Argimón *et al.* 2016).

Overall, four STs: ST-188, ST-5, ST-151 and ST-425, were limited to one farm whilst the remaining, including ST-1, were found across multiple farms. On farms from which multiple isolates were sampled, 50.0% (6/12) had two or more STs, however in most cases, the different STs were isolated from different study animals. The exception was one animal with sub-clinical mastitis in which two STs were isolated (ST-1247 and ST-1) from different teats. Across the isolates, seven STs were identified from sub-clinical samples whereas four STs were identified from cases of clinical mastitis. The STs from clinical cases were ST-1, ST-97, ST-151 and ST-705 (Table 7.1).

The screening of contigs through ResFinder identified 14 AMR genes (Appendix E, Table E3) associated with a range of drug classes and phenotypes (Table 7.2). Figure 7.3 shows how the profile of these genes vary between the isolates with some genes (*dfrC*, *fusC*, *mecA* and *mecR1*) only found to be present in a single isolate whilst one gene, *tet(38)*, was present in all 57 isolates. Screening through the Virulence Factors Database identified 76 virulence genes, 55 (72.4%) of which appeared across 100% of the isolates (Appendix E, Table E4). The profiles for those genes that varied in prevalence across the isolates can be seen in Figure 7.4, highlighting genes uncommon to all the isolates, including those only found in a single isolate (*chp*, *sea* and *selk*). In the Scoary analysis three STs: ST-1, ST-97 and ST-188, were found to be associated with a number of genes based on our criteria for significance; that is, having an odds ratio greater than one, a specificity greater than 95%, Benjamini-Hochberg corrected *p-value* below 0.05, and not being identified as either a hypothetical or putative protein.

Overall, 21 genes were identified by Scoary as having a significant association with the 35 ST-1 bovine isolates (Table 7.3). Out of these genes, 85.7% (18/21) were found only in ST-1 isolates with 47.6% (10/21) found in all 35 ST-1 isolates. The GO terms for these genes revealed that many play a role in DNA replication and modification mechanisms

Table 7.2. Drug classes influenced by the 14 resistance genes identified in the 57 *Staphylococcus aureus* isolates and their common uses in the New Zealand dairy industry. Resistance genes have been grouped by their gene family.

Gene(s)	Drug Class	Common use in New Zealand dairy industry
<i>ant(9)-Ia</i>	Aminoglycosides	Intra-mammary antimicrobials for the treatment of mastitis in lactating cows e.g. Neomycin, Lincomycin and Streptomycin
<i>blaI</i> <i>blaPC1</i> <i>blaR1</i> <i>blaZ</i> <i>mecA</i> <i>mecR1</i>	β -lactams ^a	Broad range antimicrobials used to treat a range of intra-mammary, intra-uterine and systemic infections e.g. Penicillin, Amoxicillin, Cloxacillin, Penethamate, Cefuroxime, Cephalexin, Cephalonium, Cefapirin, Ceftiofur and Cefquinome
<i>dfrC</i>	Diaminopyrimidines	Used to treat a broad variety of bacterial infections in humans, particularly urinary tract infections with limited use in cattle, with the exception of Trimethoprim, which is commonly used in combination with sulfa drugs so therefore is a fairly common treatment (e.g., Amphoprim® or Tribissen®) for enteric or respiratory tract disease in cattle
<i>erm(A)</i>	Streptogramins, Lincosamides and Macrolides	Antimicrobials used in the treatment of various systemic and localised bacterial infections including mastitis, respiratory infection, metritis and foot-rot, although Tilmicosin and Tulathromycin have a very long milk WHP so hence are not used in lactating cattle, and rarely on dairy farms. Erythromycin is also no longer used in cattle in New Zealand.
<i>fosD</i>	Fosfomycin	Used to treat a broad variety of bacterial infections in humans, particularly urinary tract infections but it is not registered for animal use in New Zealand
<i>fusC</i>	Fusidic acid	Fusidic acid is not registered for cattle use in New Zealand but has registration for use in dogs
<i>qacA</i> <i>qacB</i>	Fluoroquinolones	Injectable antimicrobials used in a range of treatment including <i>E. coli</i> and <i>Pseudomonas</i> mastitis, osteomyelitis and respiratory infections
<i>tet(38)</i>	Tetracyclines	Antimicrobial used in broad-spectrum treatment of local and systemic infections particularly uterine infections and other soft tissue infections in cattle

^a β -lactam antibiotics include Penicillins, Cephalosporins, Cephamycins, Monobactams and Carbapenems

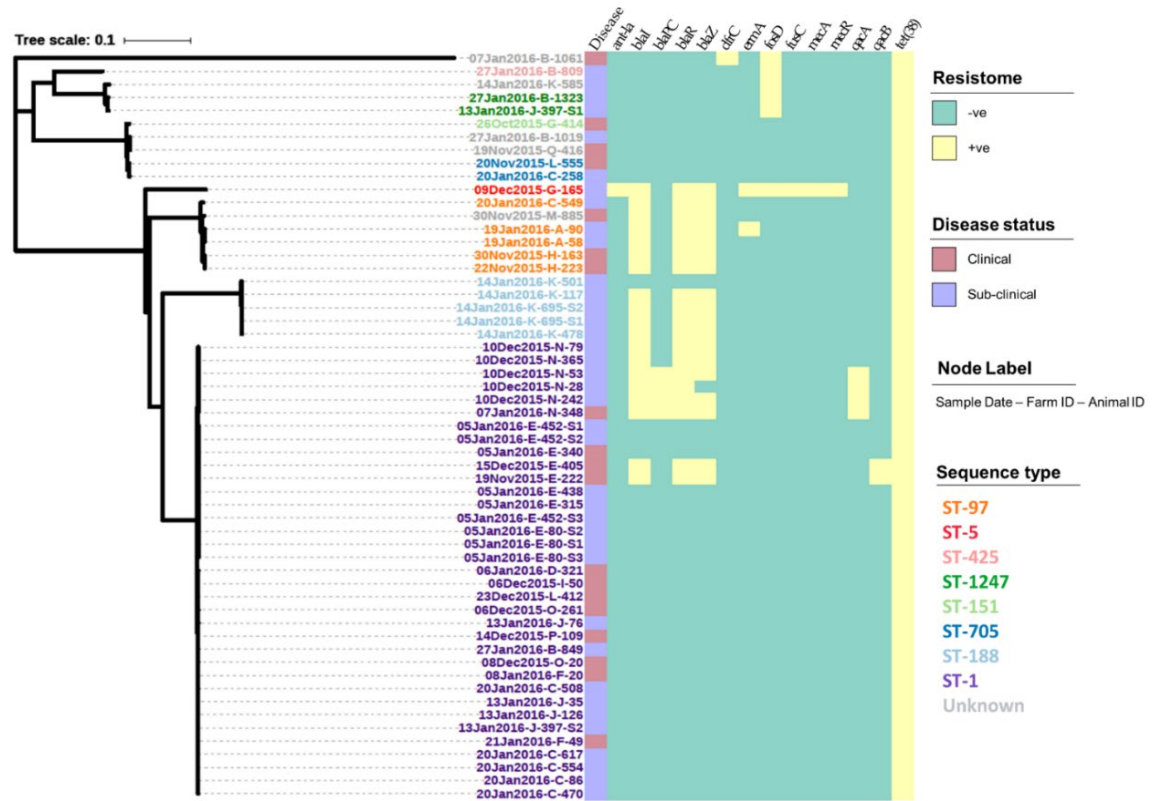


Figure 7.3. Resistance gene profiles of 57 *S. aureus* isolates presented alongside a maximum-likelihood tree generated from the core single nucleotide polymorphisms. Isolate IDs identify the date the sample was collected (dd/mmm/yyyy), the farm from which it was collected from (A-Q) as indicated in Figure 7.2, and the animal ID number (####).

whilst only three; *agrB*, *entH* and *flr*, have been linked to increased virulence and pathogenesis (Table 7.3). For the Scoary analysis, examining gene presence-absence in isolates sampled from clinical and sub-clinical mastitis cases, many candidate genes were identified; however, for all the genes the pairwise *p-value* was greater than 0.05 suggesting that after consideration of the confounding by lineages, no genes were significantly associated with either clinical or sub-clinical phenotypes. To view the complete Scoary results readers are guided to the additional file that is available in the following GitHub repository https://github.com/SSGreening/NZ_S.aureus

Figure 7.2, and the animal ID number (###).

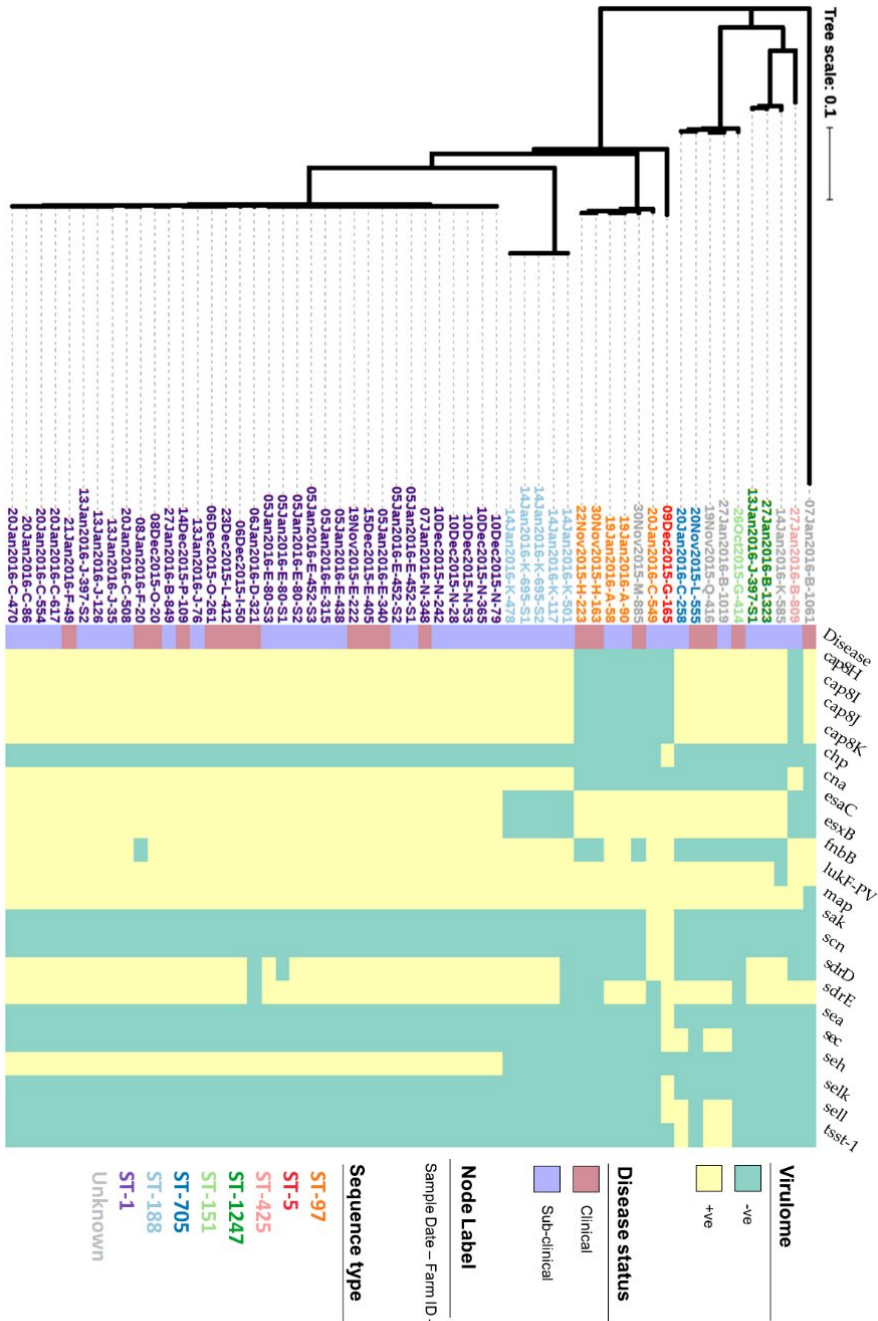


Table 7.3. Scoary result summary showing the genes found to be significantly associated with *Staphylococcus aureus* sequence type (ST)-1 and the Gene Ontology (GO) terms indicating either the biological processes or molecular functions associated with the gene products. Significant genes were those that had an odds ratio greater than 1, a specificity greater than 95%, a Benjamini-Hochberg corrected *p-value* below 0.05 and were not annotated as either a “hypothetical” or “putative” protein.

Gene	GO terms	Number of isolates gene present (%)	
		ST-1 (n = 35)	Other STs (n = 22)
<i>agrB</i>	Quorum sensing, pathogenesis and, peptidase activity	35 (100)	0
<i>entH</i>	Virulence, metal ion binding and, toxin activity	35 (100)	0
<i>flr</i>	Pathogenesis and, signal peptide	35 (100)	0
<i>catE-2</i>		35 (100)	1 (4.5)
<i>gltR</i>	Transcription regulation and, DNA-binding	35 (100)	1 (4.5)
<i>yofA</i>		35 (100)	0
<i>gdmA</i>	Cytolysis and, signalling receptor binding	35 (100)	0
<i>nisC</i>	Maturation of the lantibiotic	35 (100)	0
<i>repE</i>		35 (100)	0
<i>repN</i>	DNA replication initiation and, DNA-binding	35 (100)	0
<i>group-2156</i>		35 (100)	0
<i>group-2167</i>	Signal peptide	35 (100)	0
<i>ssbA-1</i>	DNA replication, repair and recombination and,	23 (65.7)	0
<i>ssbA-2</i>	Single stranded DNA binding	14 (40.0)	1 (4.5)
<i>dnaC-2</i>	DNA replication, synthesis of RNA primer, ATP binding, DNA binding and, DNA helicase activity	21 (60.0)	0
<i>brnQ-3</i>	Branched-chain amino acid transmembrane transporter activity	16 (45.7)	0
<i>dut</i>	dUMP biosynthetic process, dUTP diphosphatase activity and, magnesium ion binding	14 (40.0)	0

Table 7.3 continues next page

Table 7.3 continued

Gene	GO terms	Number of isolates gene present (%)	
		ST-1 (n = 35)	Other STs (n = 22)
<i>bcgIA</i>	DNA modification, DNA-binding, hydrolase activity	14 (40.0)	0
<i>bcgIB</i>	and, N-methyltransferase activity	14 (40.0)	0
<i>hin</i>	DNA integration, DNA-binding and, recombinase activity	10 (28.6)	0
<i>cna</i>	Pathogenesis, cell adhesion and, collagen binding	9 (25.7)	0

7.4.2. Antimicrobial sensitivity

Out of the six antimicrobials tested, isolates only showed full resistance to penicillin and some intermediate resistance to oxacillin (Figure 7.5). For penicillin 36.0% (18/50) showed resistance of which the majority had the *blaI* (17/18, 94.4%), *blaR1* (17/18, 94.4%), *blaZ* (16/18, 88.9%), and *tet(38)* (18/18, 100%) gene present whilst a small number also had the *blaPC1* (4/18, 22.2%), *erm(A)* (1/18, 5.6%), *qacA* (4/18, 22.2%), and *qacB* (2/18, 11.1%) gene present (Table 7.4). Note here that not all these genes are associated with penicillin resistance; for example, the *tet(38)* gene is instead known to be associated with tetracycline resistance; although in this case the *tet(38)* gene did not correlate with any tetracycline resistance. For oxacillin 6.0% (3/50) showed intermediate resistance, with all three isolates having all the *bla* genes present (*blaI*, *blaR1*, *blaZ*, and *blaPC1*) and the *qacA* gene (Table 7.4). For the full test results including the zone range for each antimicrobial, readers are directed to Appendix E (Table E5).

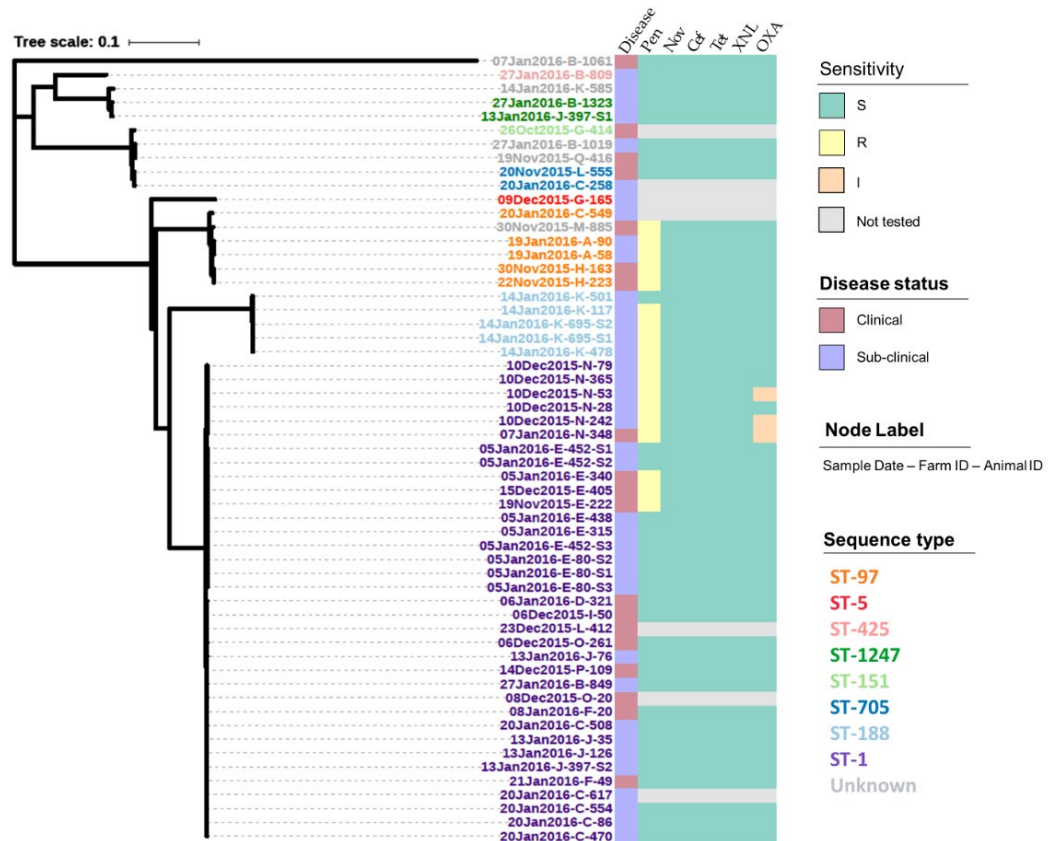


Figure 7.5. Antimicrobial sensitivity profiles of 50 *S. aureus* isolates presented alongside a maximum-likelihood tree generated from the core single nucleotide polymorphisms. Sensitivity was determined using a zone diffusion test following the procedures provided by the Clinical and Laboratory Standards Institute (CLSI) for penicillin (Pen), novobiocin (Nov), ceftiofur (Cef), tetracycline (Tet), ceftiofur (XNL), and oxacillin (OXA); with isolates being declared sensitive (S), intermediate (I) or resistant (R), based on CLSI recommendations. Isolate IDs identify the date the sample was collected (dd/mm/yyyy), the farm from which it was collected from (A-Q) as indicated in Figure 7.2, and the animal ID number (####).

Table 7.4. The presence-absence of 14 resistance genes and the antimicrobial sensitivity across 50 *S. aureus* isolates. Antimicrobial tested zone diffusion include penicillin (Pen), novobiocin (Nov), cefoxitin (Cef), tetracycline (Tet), ceftiofur (XNL), and oxacillin (OXA); with isolates being declared sensitive (S) or resistant (R) (including those also identified as intermediate), based on CLSI recommendations.

Gene presence/absence (number of isolates with gene)	Pen		Nov		Cef		Tet		XNL		OXA	
	R	S	R	S	R	S	R	S	R	S	R ^a	S
<i>ant(9)-ia</i> (n = 0)	Absent Present	18 -	32 -	0 -	50 -	0 -	0 -	50 -	0 -	50 -	3 -	47 -
<i>blaI</i> (n = 17)	Absent Present	1 17	32 0	0 0	33 17	0 0	33 17	0 0	33 17	0 0	33 17	0 3
<i>blaPC1</i> (n = 4)	Absent Present	14 4	32 0	0 0	46 4	0 0	46 4	0 0	46 4	0 0	46 4	0 3
<i>blaR1</i> (n = 17)	Absent Present	1 17	32 0	0 0	33 17	0 0	33 17	0 0	33 17	0 0	33 17	0 3
<i>blaZ</i> (n = 16)	Absent Present	2 16	32 0	0 0	34 16	0 0	34 16	0 0	34 16	0 0	46 4	0 3
<i>dfnC</i> (n = 1)	Absent Present	18 0	31 1	0 0	49 1	0 0	49 1	0 0	49 1	0 0	33 17	0 3

Table 7.4 continues next page.

Table 7.4 continued.

Gene presence/absence (number of isolates with gene)	Pen		Nov		Cef		Tet		XNL		OXA	
	R	S	R	S	R	S	R	S	R	S	R ^a	S
<i>erm(A)</i>	17	32	0	49	0	49	0	49	0	34	3	46
(n = 1)	1	0	0	1	0	1	0	1	0	16	0	1
<i>fosD</i>	18	27	0	45	0	45	0	45	0	49	3	42
(n = 5)	0	5	0	5	0	5	0	5	0	1	0	5
<i>fusC</i>	18	32	0	50	0	50	0	50	0	50	3	47
(n = 0)	-	-	-	-	-	-	-	-	-	-	-	-
<i>mecA</i>	18	32	0	50	0	50	0	50	0	50	3	47
(n = 0)	-	-	-	-	-	-	-	-	-	-	-	-
<i>mecR1</i>	18	32	0	50	0	50	0	50	0	50	3	47
(n = 0)	-	-	-	-	-	-	-	-	-	-	-	-
<i>qacA</i>	14	32	0	46	0	46	0	46	0	46	0	46
(n = 4)	4	0	0	4	0	4	0	4	0	4	3	1
<i>qacB</i>	16	32	0	48	0	48	0	48	0	48	3	45
(n = 2)	2	0	0	2	0	2	0	2	0	2	0	2
<i>tet(38)</i>	0	0	0	0	0	0	0	0	0	0	0	0
(n = 50)	18	32	0	50	0	50	0	50	0	50	3	47

-no isolates tested had gene present

^a all isolates in R (*i.e.*, resistant) were classified as intermediate

7.4.3. Comparative genomics: *S. aureus* within New Zealand

The maximum-likelihood phylogeny of the 57 *S. aureus* isolates from this study and the 59 *S. aureus* isolates from the previous study by Grinberg and colleagues (2010) is shown in Figure 7.6 (Grinberg *et al.* 2017). Overall, 42,929 polymorphic sites were identified between the 116 isolates containing 22,484 core SNPs. In total 15 STs were identified (Table 7.5) of which, only three: ST-5, ST-188 and ST-1, contained isolates sampled from all the species; that is cattle, humans, dogs and cats. Appendix E, Figure E2 provides a higher resolution phylogeny of the ST-1 cluster containing 35 isolates sampled from bovine milk samples, nine isolates sampled from humans, two isolates sampled from pet cats and five isolates sampled from pet dogs. Other STs were limited to bovine-derived isolates including ST-97, ST-151, ST-425, ST-705 and ST-1247 whilst the remaining included isolates from both humans and pets. Overall, there was a higher prevalence of ST-1 isolates in bovine-derived isolates in comparison to isolates sampled from either humans, cats or dogs but a lower prevalence of both ST-188 and ST-5 (Table 7.5)

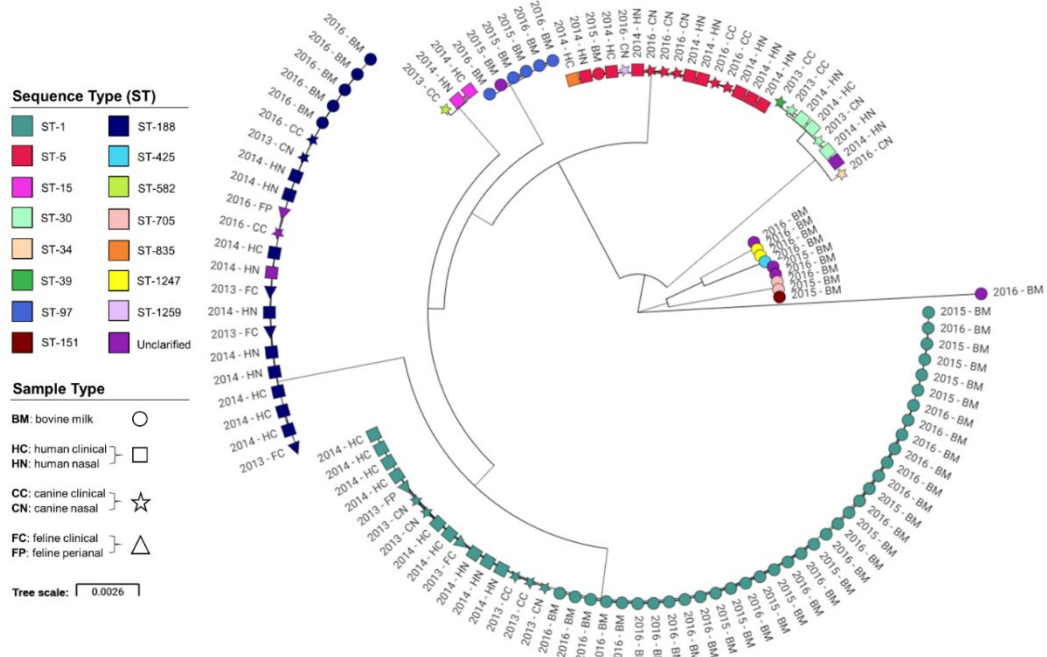


Figure 7.6. Maximum-likelihood phylogeny generated from core single nucleotide polymorphisms across 57 *S. aureus* isolates from this study (*i.e.*, New Zealand dairy cattle) and 59 *S. aureus* isolates from a previous study by Grinberg and colleagues (Grinberg *et al.* 2010). Isolate IDs identify the year the sample was collected (yyyy) and the species the sample was taken from. The colour of the terminal tree node indicates the sequence type identified by seven MLST genes whilst the shape further identifies the species and site from which the isolate it was collected. BM: bovine milk, HC: human clinical case, HN: human nasal colonisation, CC: canine clinical case, CN: canine nasal colonisation, FC: feline clinical case and FP: feline perianal colonisation.

Table 7.5. The dominant *Staphylococcus aureus* 7 gene multilocus sequence types (MLST) observed in our study looking at 57 *S. aureus* isolates sampled from bovine milk samples, and previous studies conducted by Grinberg *et al.* (2017) (Study 1) including 59 *S. aureus* isolates sampled from humans (n = 34) and household pets (cats = 6, dogs = 19) from across New Zealand and Heffernan *et al.* (2015) (Study 2) including 1255 methicillin-susceptible *S. aureus* (MSSA) isolates sampled from humans from across New Zealand. Note the MLST for five of the isolates in the current study (5/57, 8.8%) could be detected as well as three in study 1 (3/59, 5.1%). For study 2 only the seven dominant MSSA clones have been identified therefore, the column percentages do not equal 100%.

7 gene MLST types	Source	Prevalence (% of all isolates)		
		Current study	Study 1	Study 2
ST-1	Bovine	35 (61.4)	-	-
	Human	-	9 (15.3)	192 (15.3)
	Pets	-	7 (11.9)	-
ST-188	Bovine	5 (8.8)	-	-
	Human	-	9 (15.3)	128 (10.2)
	Pets	-	5 (8.5)	-
ST-5 (and ST-835) ^a	Bovine	1 (1.8)	-	-
	Human	-	9 (15.3)	108 (8.6)
	Pets	-	5 (8.5)	-
ST-705	Bovine	2 (3.5)	-	-
	Human	-	-	-
	Pets	-	-	-
ST-1247	Bovine	2 (3.5)	-	-
	Human	-	-	-
	Pets	-	-	-
ST-97	Bovine	5 (8.8)	-	-
	Human	-	-	-
	Pets	-	-	-
ST-151	Bovine	1 (1.8)	-	-
	Human	-	-	-
	Pets	-	-	-
ST-425	Bovine	1 (1.8)	-	-
	Human	-	-	-
	Pets	-	-	-

Table 7.5 continued

Table 7.5 continued

wgMLST types	Source	Prevalence (% of all isolates)		
		Current study	Study 1	Study 2
ST-15	Bovine		-	-
	Human		2 (3.4)	35 (2.8)
	Pets		-	-
ST-30	Bovine		-	-
	Human		3 (5.1)	24 (1.9)
	Pets		2 (3.4)	-
ST-34	Bovine		-	
	Human		-	
	Pets		1 (1.7)	
ST-39	Bovine		-	
	Human		-	
	Pets		1 (1.7)	
ST-582	Bovine		-	
	Human		-	
	Pets		1 (1.7)	
ST-1259	Bovine		-	
	Human		-	
	Pets		1 (1.7)	
ST-121 (and ST-2276) ^b	Bovine			-
	Human			73 (5.8)
	Pets			-

^a Combined prevalence of ST-5 and ST-835, a single-locus variant of ST-5, and therefore belonging to the same MLST clonal complex 5 (CC5)

^b Combined prevalence of ST-121 and ST-2276, a double-locus variant of ST-121, and therefore belonging to the same MLST clonal complex 121 (CC121)

7.4.4. Comparative genomics: *S. aureus* ST-1 globally

Overall, 9,897 WGS *S. aureus* assemblies were available from GenBank as of March 2019. After wgMLST it was revealed that 25.8% (2,550/9,897) belonged to one of the eight STs identified for this study of which 4.4% (103/2,550) were ST-1; the predominant ST in this study (Appendix E, Figure E1). These included isolates sampled from across 19 countries (Figure 7.7) between the years 1998 to 2017. From these isolates, 4.9% (5/103) were collected from milk samples from cattle originating in the United Kingdom, Australia, Brazil and Italy compared to 93.2% (96/103) that were collected from humans including blood cultures, nasal swabs, pharyngeal swabs, skin lesion and wounds, sputum samples,

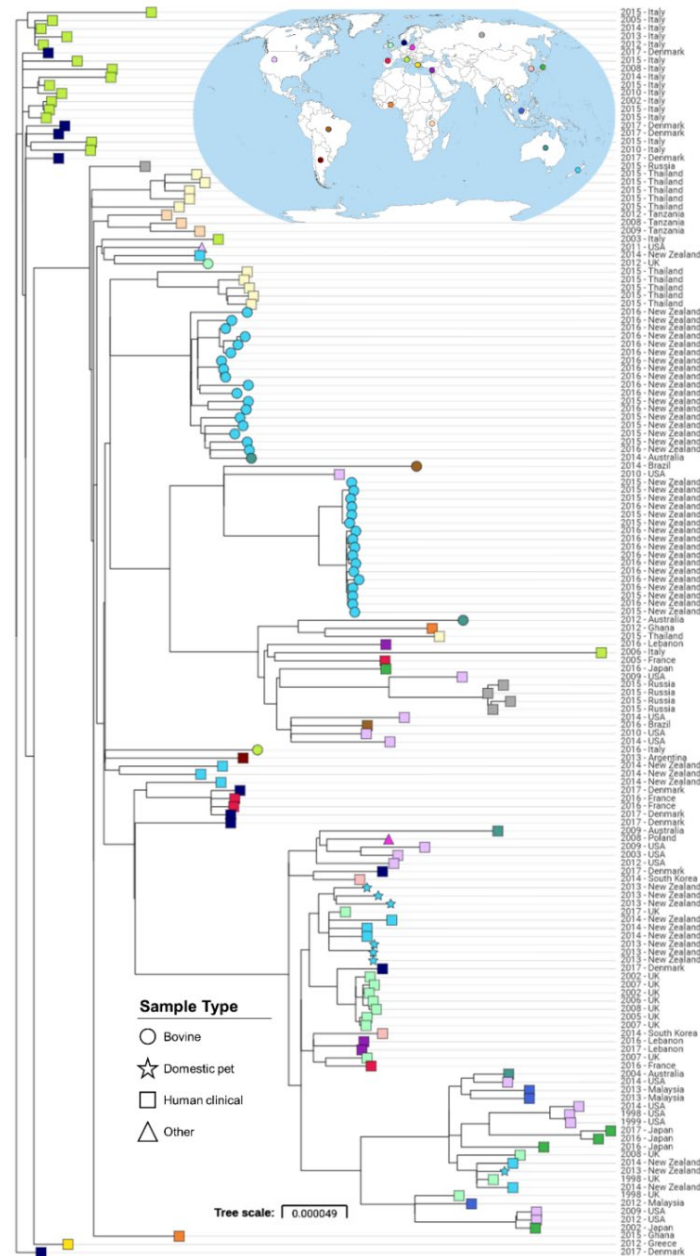


Figure 7.7. (a) Maximum-likelihood phylogeny generated from core single nucleotide polymorphisms across 35 *S. aureus* ST-1 isolates from this study (*i.e.*, New Zealand dairy cattle) and the 103 *S. aureus* ST-1 isolates available in Genbank as of March 2019. Isolate IDs identify the year the sample was collected (yyyy) and the country from which it was sampled from whilst the shape of the terminal node indicates the species it was sampled from (circle = cattle, square = human, star = domesticated pets and, triangle = all other samples). (b) Map showing the countries from which the isolates were sampled from with points sitting on the centroid position. Point shape indicates the dominate species that has been sampled in the country.

ear swabs and, broncho- and tracheal-aspirations. The two (1.9%) remaining isolates were sourced from broiler chickens in Poland and the assembly production room of a bakery in the United States. The maximum-likelihood phylogeny of the 35 *S. aureus* ST-1 isolates from this study, the 16 *S. aureus* ST-1 isolates from the previous study by Grinberg and colleagues (2010) and the 103 *S. aureus* ST-1 isolates described above is shown in Figure 7.7 (Grinberg *et al.* 2017). They cluster according to 7,072 polymorphic sites and 7,010 core SNPs identified between the 154 ST-1 isolates. This phylogeny shows a monophyletic group, which has been highlighted in Figure 7.7, containing 18 New Zealand bovine isolates and one Australian isolate also from a bovine raw milk sample. The addition of the remaining 17 New Zealand bovine isolates creates a polyphyletic group including one human isolate from the USA and a bovine isolate from Brazil. These 17 New Zealand isolates also form a second more diverse monophyletic group containing not only another Australian bovine isolate but also human isolates Thailand, Brazil, USA, Ghana, Lebanon, Italy, France, Japan and Russia. The New Zealand ST-1 isolates samples from both humans and pets can be found throughout the phylogeny, clustering with isolates collected from across all the study countries.

The Scoary analysis found a number of genes to be overrepresented in isolates sampled from 4/5 of the geographical regions, excluding South America. However, no genes were considered of interest in those isolates sampled from North America and Africa. In isolates sampled from Asia, 25 genes were found to be overrepresented, including several genes associated with Pathogenicity Island proteins, whilst 26 genes were found to be overrepresented in isolated sampled from Europe, including the resistance genes *mecA* and *ermA*, and 51 genes overrepresented in isolates sampled from across Oceania, including a number of genes associated with phage-related proteins. However, all the genes found to be significantly associated with isolates sampled from Asia, Europe and Oceania were only present in a small proportion of the total isolates from that region.

Further Scoary results looking at ST-1 gene presence-absence and host species, found the presence of 93 different genes to be significantly overrepresented in bovine-derived isolates of which 54.8% (51/93) were limited to only the bovine-derived isolates although none were found across all of the bovine isolates. Most interestingly a small number of these genes (26.9%; 25/93) were identified as phage-related proteins with many also being identified in the human isolates but not in the isolates sampled from pets. Out of the remaining genes, many are thought to act as transcriptional regulators with two genes; *lukD* and *lukE*, associated with increased virulence via toxin activity and cytolysis found in 37.5% (15/40) of the isolates. Overall, 72 genes were found to be significantly overrepresented with the human-derived isolates of which 43.1% (31/72) were limited to only the human isolates. Out of these genes, four have been associated with enterotoxin production and increased virulence, with the two most prevalent; *sem* and *sei*, found in over 55% of the human isolates. In addition to these virulence genes, three genes associated with resistance; *mecR1*, *mecA* and *ermA* were also found to be overrepresented in the human isolate with a presence in 41.0% (43/105), 32.4% (34/105), and 29.5% (31/105) of the isolates respectively. Lastly, 32 genes were found to be significantly overrepresented with isolates sampled from pets of which 43.1% (23/32) were limited to only the pet isolates, however, no GO terms could be found for any of these genes. To view the complete Scoary results readers are guided to the additional file that is available in the following GitHub repository https://github.com/SSGreening/NZ_S.aureus.

7.4.5. Cattle trade network

The LIC-MINDA database included 150,315 movement records between 45,641 different farms. In total, 82.4% (14/17) of the farms in this study had recorded a live animal movement in the LIC-MINDA database, suggesting that three of the herds were closed herds with no on- or off-farm live animal movements prior to 2011, or more likely a lack of data recording. Network statistics have been presented in Table 7.6. To summarise, there seemed to be lack of cohesiveness between all the farms in the network with only a

Table 7.6. Summary network statistic for all on- and off-farm live animal movements recorded in the Livestock Improvement Corporation (LIC) MINDA database over a 26-year period from 1985-2011.

Network property	Network measure	Value
Network size	<i>Number of nodes</i>	45,641
	<i>Number of edges</i>	150,315
	<i>Network diameter</i>	21
	<i>Average path length</i>	6.86
Centrality measures	<i>Average in-degree (min-max)</i>	3.29 (0 - 227)
	<i>Average out-degree (min-max)</i>	3.29 (0 - 110)
	<i>Average betweenness (min-max)</i>	42356 (0 - 1.8X10 ⁷)
Cohesion measures	<i>Network density</i>	7.21 x 10 ⁻⁵
	<i>Clustering coefficient</i>	0.029
	<i>Giant Strongly Connected Component (GSCC)</i>	19,965
	<i>Giant Weakly Connected Component (GWCC)</i>	31,681
	<i>Reciprocity</i>	0.078
	<i>Fragmentation</i>	0.809
	<i>Assortativity</i>	0.002

small proportion of all possible connections being made resulting in a highly fragmented network structure. However, when considering all live animal movements, 43.7% (19,965/45,641) of farms form the giant strongly connected component indicating some level of connectivity. In total, the fast-greedy community analysis identified 5,740 different communities; the largest of which contained 152 farms, however, despite the large number of communities 64.7% (11/17) of the study farms belonged to a single community. Overall, all the study farms were grouped into six different communities including the three farms with no animal movements records as isolated communities.

7.4.6. Antimicrobial usage

The antimicrobial sales data showed a large amount of heterogeneity in the purchasing of antimicrobials between the study farms. The DD was calculated per cow per year to account for differences in the herd size which ranged from 96 to 950 animals (mean = 512). Overall the DD ranged from 0.08 to 3.02 per cow per year (mean = 1.57) however this varied by treatment method, with intra-mammary treatments having the highest DD

ranging from 0 to 2.76 (mean = 1.41) followed by injectable (mean = 0.33) and intra-uterine treatments (mean = 0.05). To investigate if larger herds used more antimicrobials, the SCC was calculated, however no association was found between herd size and DD (SCC = 0.27, $P = 0.263$) (Appendix E, Figure E3). To investigate if an increase in AMU was associated with the total number of resistance genes present a further SCC was calculated however no association was found (SCC -0.32, $p = 0.204$) (Appendix E, Figure E4). When considering only dry cow therapy, five treatment methods were reported. The use of Bovaclox™ was reported by 23.5% (4/17), Cepravin® by 35.3% (6/17), Dryclox® by 17.6% (3/17), Orbenin® by 47.1% (8/17), and Teatseal® by 58.9% (10/17).

Scoary analysis found no association between any of the dry cow therapy treatments and any of the resistance genes identified through ResFinder (Zankari, *et al.* 2012) and CARD (Jia *et al.* 2017). In fact, no genes were found to be significantly associated with treatment when considering the pairwise *p-values* however, when making no causal inference one gene; *yezG*, known for its role in the YeeF-YezG toxin-antitoxin module (Christensen *et al.* 2016), was found to be significantly overrepresented (Benjamini-Hochberg *p-value* = 0.048) in isolates sampled from reporting the use of Orbenin®. A number of genes were also found to be overrepresented in isolates sampled from farms reporting the use of Bovaclox™ of which 83.8% (31/37) were found in all the isolates positive for this trait including the *merR1* gene which is thought to be the principal regulatory gene that controls the expression of the *merA* operon responsible for mercury resistance in bacteria (Stapleton *et al.* 2004; Ojo *et al.* 2004). The *norB4* gene was also found in all the isolates sampled from farms that had reported using Bovaclox™, and although this gene was not identified by either ResFinder or CARD, it has previously been associated with ciprofloxacin-resistant *S. aureus* (Kwak *et al.* 2013). To view the complete Scoary results readers are guided to the additional file that is available in the following GitHub repository https://github.com/SSGreening/NZ_S.aureus.

7.4.7. Permutational multivariate analysis of variance

The univariate PERMANOVA models showed that there was a significant difference in the core SNP profiles between the different farms ($p = 0.0046$), townships ($p = 0.0004$) and network communities ($p = 0.0001$) (Appendix E, Table E6) with differences between farms accounting for more of the variation between the genetic profiles than differences between townships or communities. When comparing two mixed-design PERMANOVA models, with farms nested either within township or community, community effects (Table 7.7) were stronger ($p = 0.0009$) than township ($p = 0.0146$). However, care must be taken when interpreting the multivariate PERMANOVA models as the Chi-squared test showed a strong association between all the predictor variables (Appendix E, Table E7) suggesting collinearity. MDS ordination plots based on the uncorrected p -measure identified one isolate relatively distant from the other isolates. However, when this isolate was removed, three clusters emerged from the core SNP profiles with isolates more closely related to others within the same cluster compared to isolates in neighbouring a cluster. From the MDS ordination plots alone it is difficult to determine if isolates within these clusters are also grouped by farm, community or township despite the associations found in the univariate PERMANOVA models (Appendix E, Figure E5).

Table 7.7. Multivariate Permutational Multivariate Analysis of Variance (PERMANOVA) model comparing the p -dissimilarity measure between the core genome single nucleotide polymorphism (SNP) profiles of 57 *Staphylococcus aureus* isolates by network community ($n = 6$) and farm (nested within community) ($n = 17$) with p -values for each test obtained using 9999 unrestricted permutations. “SS” provides the sum of squares, “MS” the mean squares, and “ df ” the degrees of freedom for each test.

	<i>df</i>	SS	MS	Pseudo- <i>F</i>	<i>p-value</i>
Community^a	5	0.77	0.15	0.04	0.0009
Farm	11	0.39	0.04	0.40	0.4684
Residuals	40	1.40	0.04		
Total	56	2.56			

7.5. Discussion

Overall, this study identified eight *S. aureus* sequence types amongst 57 isolates derived from bovine milk samples with a number of STs identified from isolates collected both within the same herd and from the same animal. Altogether, 14 genes associated with antimicrobial resistance and 76 genes associated with virulence factors were identified with very little variation in the gene profiles both within a single ST and between STs except for one ST-5 isolate. Furthermore, despite a number of resistance genes being identified very few of the isolates showed evidence of resistance to any of the antimicrobials tested with the exception of the *bla* genes and penicillin resistance. This highlights the importance of performing phenotypic tests for antimicrobial susceptibility with many weaknesses identified in genome-based predictions (Courvalin, 2005; Gordon et al. 2014). Overall, the low level of diversity may be a result of a very strong farm effect, with isolates from the same farm, belonging to the same ST, having very similar genetic profiles. In addition to farm-level effects, the genetic relatedness of isolates was found to be associated with both trading community and township; supporting previous evidence that the movement of live animals may be an important risk factor for the spread of *S. aureus*. However, given that some closely related isolates were not connected in the animal trade network, it suggests that other between farm contacts, such as the shared milking parlours or neighbouring paddocks, may also be playing a role in the spread of *S. aureus* (Álvarez et al. 2011). Given the likelihood that some of the isolates may also have human origin, it would also be important to consider other human contacts with animals in local communities such as the movement of personnel between farms.

When comparing the current study bovine-derived isolates with additional *S. aureus* isolates collected from domestic pets and humans across New Zealand, only three STs were identified across all species; ST-1 (the predominant ST in the bovine isolates), ST-5, and ST-188. These STs are thought to be derived from lineages associated with human infection whilst historically found to be uncommon among bovine isolates in comparison

to human MRSA and MSSA isolates (Hata *et al.* 2010). This result is consistent with the growing emergence of human-associated STs as causative agents of bovine mastitis that has been reported in a number of studies worldwide (Sakwinska *et al.* 2011; Bar-Gal *et al.* 2015); supporting evidence of *S. aureus* as a zoonotic pathogen. Nevertheless, without collecting further isolates over a greater geographical distribution and longer timeframe, it is not possible to identify if these human-derived STs are persisting within the livestock populations.

In the global comparison of ST-1 isolates, the bovine isolates from the current study formed two genetic clusters within a monophyletic group whereas, the isolates from both domesticated pets and humans in New Zealand were spread throughout several clusters including isolates collected from many other countries. This clustering of New Zealand bovine ST-1 isolates may also be an artefact resulting from the current studies small sample size that is limited to one geographical region. The restriction may also explain the low prevalence of ST-97 and ST-705, the two lineages thought to be otherwise dominant worldwide from isolates derived from bovine milk (Hata *et al.* 2010). However, it may also be possible that New Zealand's strict border control and import regulations for live animals have limited the introduction of *S. aureus* from other countries with the last live cattle import from Australia into New Zealand dating back to 2008, and live imports from other countries ceasing in the early '90s; resulting in genetically distinct clusters. Nevertheless, the presence of a closely related isolate originating from Australia does suggest possible transmission within Australasia which could be either from a historic cattle import or derived from human isolates that have subsequently adapted to cattle, although without further genetic comparison between New Zealand and Australian isolates it is not possible to infer any transmission dynamics. Further investigation could be conducted to try and trace any relevant epidemiological links that may have resulted in transmission, such as any historic cattle trades between the two countries, although it would be hard to define a time period in which to restrict the search

as the small number of isolates characterised in the current study over a short time period makes it difficult to trace the time to the most recent common ancestor (MRCA); a common technique used in phylogenetics to study the shared evolutionary history of isolates and infer transmission. The close relationship between the Australian and New Zealand isolates may not be explained by cattle trade but could instead be the result of human movement, particularly of those who work in close contact with livestock such as farm workers or veterinarians, which would be harder to trace.

In addition to the small sample size, there are many other limitations in the study that must be taken into account when interpreting the results. For example, in the genomic analysis an ST-1 reference genome was used to create the SNP alignment and identify differences between the isolates; a method that is widely used but not without its limitations as the accuracy of SNP calling can be reduced when applied to relatively diverse genomes (Olson *et al.* 2015; Bush *et al.* 2019). For this study, this means that there is a higher discriminatory power when mapping the ST-1 isolates back to the reference in comparison to the other STs which are more distantly related and may even have entire regions that cannot be aligned with the reference sequence. This could result in alignments errors and biases when SNP calling (Nielsen *et al.* 2011). However, this drawback is not present in the *wgMLST* analysis which relies on *de novo* assembly, an alternative to mapping in which no reference genome is used (Wilson 2012), and has been shown to provide a significantly higher resolution and epidemiological concordance when compared with reference-based methods (Carleton and Gerner-Smidt, 2016; Nadon *et al.* 2017).

In addition to limitations with the genomic analysis, there are also many restrictions in the network analysis performed in this study. In order to construct the network graphs movement records from the LIC database were used however, it is likely that some movements have not been recorded due to human error. Movements may also be lost due

to the turnover of farms for examples, over 45,000 different farms were identified in the LIC database however, DairyNZ recorded just over 12,000 herds in 2017/18 (DairyNZ, 2018). This makes it clear that herd identifiers may be being lost or changed over time, making it difficult to historically trace the movements between farms. In addition to these limitations in the data, the extract used in this study only showed movements up until 2011 whilst the *S. aureus* isolates were collected in the 2015/16 lactation. Therefore, all the movement between the study herds that occurred between 2012 to 2016 would not be represented in the contact networks. Furthermore, these records do not consider other contacts that could be important for transmission, such as the movement of farm visitors, resulting in further missing link data between farms that will inherently change the structure of the final network and hinder downstream network analysis unreliable. However, this study is most interested in network communities which have been shown to withstand the effects of missing data better than other network statistics (Kossinets, 2006; Yan and Gregory, 2011).

Despite these limitations, the gene profiles characterised in this study still allow for some comparison not only within the study population but with previous studies in order to identify changing patterns of resistance which may help to guide clinical decisions around *S. aureus* management, especially considering New Zealand's current target to completely stop the use of antibiotics for the maintenance of animal health and wellness by 2030 (Hillerton and Allison, 2015); that is, antimicrobials will not be used for preventative or metaphylactic disease therapy in animals. In light of this target, it is promising to see the small amount of diversity in the resistance gene profiles of all the bovine-derived *S. aureus* isolates with only one ST-5 isolate having *mecA* and *mecR* genes present; both of which are associated with methicillin-resistant *S. aureus* (Aires-de-Sousa, 2017). Unfortunately, this isolate did not undergo any of the sensitivity testing meaning no further conclusions can be drawn; however, being the only isolate in the study with genes associated with methicillin-resistant fits in with the current understanding that the risk of LA-MRSA

originating from cattle is minimal; with only one methicillin-resistant isolate ever having been found from a bovine source in New Zealand (Grinberg *et al.* 2008). However, this relies on the ability to extrapolate the current findings to the wider population of *S. aureus* which is difficult given that New Zealand lacks a good national survey at either the cow or bulk milk level. Nevertheless, given these findings, it appears that MRSA is relatively uncommon in cattle, in comparison to many European countries where cases have been on the rise (Spoor *et al.* 2013; Sharma *et al.* 2016), and it would suggest that the efficacy of methicillin antibiotics has been preserved.

On the other hand, all the *bla* genes associated with penicillin resistance were found present in 33.3% (19/57) of the isolates; the majority of which were identified as ST-5, ST-97 or ST-188, and showed resistance in the sensitivity test to penicillin. Both this ST profile and sensitivity pattern is consistent with previous studies; for example, in a study by van den Borne and colleagues (2010) penicillin resistance was found to be higher in ST lineages with human origin (including ST-5 and ST-188), although the majority of STs derived from bovine origin, such as ST-97 in the current study, showed little to no resistance (van den Borne *et al.* 2010). A further study by Steele and McDougall (2014), found approximately 45% of *S. aureus* isolates to be both phenotypically penicillin resistant and genotypically *blaZ* positive, and presence of this genotype/phenotype was associated with very poor bacteriological cure following antimicrobial therapy (Steele and McDougall, 2014). This high level of resistance to penicillin in comparison to all the other antimicrobial tested raises some concerns with many penicillins regarded as “critically important” (*e.g.*, ampicillin) or “highly important” (*e.g.*, cloxacillin) (WHO, 2018). However, without either further susceptibility testing or more details describing antimicrobial usage in New Zealand, it is difficult to determine which antimicrobials are most effected.

Of further concern is also the significant diversity in virulence genes that was found among only a limited number of isolates. In particular, some of these genes have been associated with food poisoning in humans namely a group of highly heat resistant superantigens called staphylococcal enterotoxins (SE) (Hennekinne *et al.* 2012). To date, there are over 20 described SEs with the majority of foodborne illnesses traced to five main serological groups; SEA, SEB, SEC, SED and SEE that are known to survive the pasteurisation process. In this study, the prevalence of SE genes was high with 70.2% (40/57) of isolates harboured at least one SE gene whilst 8.8% (5/57) had two or more. This difference is largely due to the association between SEs and STs as all ST-1 isolates, the predominant ST in this study, had the SE gene *seh* present, whilst only ST-705 and ST-5 had multiple SEs present as well as the enterotoxin-like genes *selk* and *sell* and the toxic shock syndrome gene *tsst-1*, all of which have been associated with human disease after the consumption of raw milk (Argudin *et al.* 2012; Nazari *et al.* 2014). However, overall the prevalence of SE genes in this study is comparatively lower than that previously reported by many other countries; including Italy (Riva *et al.* 2015), Germany (Zschöck *et al.* 2005), United States (Bar-Gal *et al.* 2015) and Australia (McMillan *et al.* 2016), and their presence alone is not indicative of the level of expression and toxin production in milk. Therefore, given the limitation in genome-based virulence prediction, it is important for future studies to conduct more phenotypic tests on samples that are more representatives of the wider population in order to fully assess the risk to public health, particularly regarding the consumption of raw milk.

7.6. Conclusion

By characterising the genetic population structure of mastitis-causing *S. aureus* within a limited number of New Zealand dairy herds, this study has provided evidence for the predominance of an ST formerly associated with human infection. However, despite these STs having historically higher rates of resistance, the presence of resistance genes remains low with little diversity between *S. aureus* isolates sampled from the different

farms. Lastly, by comparing the genetic relatedness of *S. aureus* isolates within and between trading communities this study provides evidence that the movement of live animals may be a risk factor for the spread of *S. aureus*, and although this highlights the importance of animal movement records in epidemiological investigations, it is also clear that using cattle-tracing data alone may not be enough to fully capture disease dynamics with further evidence of local spread.

7.7. Acknowledgements

The authors are grateful to all the participating farms as well as the team at Anexa FVC for collecting the samples and LIC for access to the MINDA data. Funding for this research was by the School of Veterinary Science at Massey University and New Zealand's Biological Heritage: National Science Challenges.

7.8. References

- Aires-de-Sousa M. Methicillin-resistant *Staphylococcus Aureus* among animals: current overview. *Clin. Microbiol. Infect.* 23, 373–80, 2017
- Álvarez LG, Webb CR, Holmes MA. A novel field-based approach to validate the use of network models for disease spread between dairy herds. *Epidemiol. Infect.* 139, 1863–1874, 2011
- Anderson MJ, Gorley RN, and Clarke KR. *PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods*. PRIMER-E Ltd., Plymouth, UK, 2008
- Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, Feil EJ, Holden MTG, Yeats CA, Grundmann H, Spratt BG, Aanensen DM. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb. Genom.* 2, e000093-104, 2016
- Argudín, MA, Mendoza MC, González-Hevia MA, Bances M, Guerra B, and Rodicio MR. Genotypes, exotoxin gene content, and antimicrobial resistance of *Staphylococcus Aureus* strains recovered from foods and food handlers. *Appl. Environ. Microbiol.* 78, 2930–35, 2012
- Bansal S, Grenfell BT, Meyers LA. When individual behaviour matters: homogeneous and network models in epidemiology. *J. R. Soc. Interf.* 4, 879–891, 2007
- Bar-Gal GK, Blum SE, Hadas L, Ehricht R, Monecke S, Leitner G. Host specificity of *Staphylococcus aureus* causing intra-mammary infections in dairy animals assessed by genotyping and virulence genes. *Vet. Microbiol.* 175, 143–54, 2015
- Becker K, Ballhausen B, Kahl B, and Köck R. The clinical impact of livestock-associated methicillin-resistant *Staphylococcus aureus* of the clonal complex 398 for humans. *Vet. Microbiol.* 200, 33-38, 2015
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Methodol.* 57, 289–300, 1995
- Biek R, O'Hare A, Wright D, Mallon T, McCormick C, Orton RJ, McDowell S, Trewby H, Skuce RA, Kao RR. Whole genome sequencing reveals local transmission patterns of *Mycobacterium bovis* in sympatric cattle and badger populations. *PLoS Pathog.* 8, e1003008, 2012

- Bogni CI, Odierno LM, Raspanti CG, Giraudo JA, Larriestra A, Lasagno MC, Ferrari MH, Ducros E, Frigerio C, Bettera SG, Frola ID, Dieser SA, Vissio C.** *War against mastitis: Current concepts on controlling bovine mastitis pathogens*. In: Science against microbial pathogens: communicating current research and technological advance, Formatex Research Centre, Badajoz, Spain, 2011.
- Bolger AM, Lohse M, Usadel B.** Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120, 2014
- Bonsaglia ECR, Silva NCC, Rossi BF, Camargo CH, Dantas STA, Langoni H, Guimarães FF, Lima FS, Fitzgerald JR, Fernandes A, Rall VLM.** Molecular epidemiology of methicillin-susceptible *Staphylococcus aureus* (MSSA) isolated from milk of cows with subclinical mastitis. *Microb. Pathog.* 124, 130–135, 2018
- van den Borne BHP, Nielen M, van Schaik G, Melchior MB, Lam TJGM, Zadoks RN.** Host adaptation of bovine *Staphylococcus aureus* seems associated with bacteriological cure after lactational antimicrobial treatment. *J. Dairy Sci.* 93, 2550–2558, 2010
- Bradley AJ.** Bovine Mastitis: An Evolving Disease. *Vet. J.* 164, 116–128, 2002
- Broeckl S, Krebs S, Varadharajan A, Straubinger RK, Blum H, Buettner M.** Investigation of intra-herd spread of *Mycobacterium caprae* in cattle by generation and use of a whole-genome sequence. *Vet. Res. Comm.* 41, 113–128, 2017
- Brynildsrud O, Bohlin J, Scheffer L, Eldholm V.** Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* 17, 2016
- Bush SJ, Foster D, Eyre DW, Clark EL, Maio ND, Shaw LP, Stoesser N, Peto TEA, Crook DW, Walker AS.** Genomic diversity affects the accuracy of bacterial SNP calling pipelines. *bioRxiv*, e653774, 2019.
- Carleton HA, Gerner-Smidt P.** Whole-genome sequencing is taking over foodborne disease surveillance. *Microbe* 11, 311–317, 2016
- Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q.** VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 33, 325–328, 2005
- Christensen GJM, Scholz CFP, Enghild J, Rohde H, Kilian M, Thürmer A, Brzuszkiewicz E, Lomholt HB, Brüggemann H.** Antagonism between *Staphylococcus epidermidis* and *Propionibacterium acnes* and its genomic basis. *BMC Genomics* 17, 2016

- Christley RM, Pinchbeck GL, Bowers RG, Clancy D, French NP, Bennett R, Turner J.** Infection in social networks: using network analysis to identify high-risk individuals. *Am. J. Epidemiol.* 162, 1024–1031, 2005
- Ciccolini M, Dahl J, Chase-Topping ME, Woolhouse MEJ.** Disease transmission on fragmented contact networks: livestock-associated Methicillin-resistant *Staphylococcus aureus* in the Danish pig-industry. *Epidemics* 4, 171–178, 2012
- Clauset A, Newman MEJ, Moore C.** Finding community structure in very large networks. *Phys. Rev. E.* 70, e066111, 2004
- Costa WLR, Ferreira J dos S, Carvalho JS, Cerqueira ES, Oliveira LC, Almeida RC de C.** Methicillin-resistant *Staphylococcus aureus* in raw meats and prepared foods in public hospitals in Salvador, Bahia, Brazil. *J. Food Sci.* 80, 147–150, 2015
- Courvalin P.** Antimicrobial drug resistance: “prediction is very difficult, especially about the future”. *Emerg. Infect. Dis.* 11, 1503-1506, 2005
- Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR.** Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* 43, e1196, 2015
- Csardi G, Nepusz T.** The igraph software package for complex network research. *InterJournal, Complex Systems*, 2006. URL: <http://igraph.org>
- Cuny C, Wieler LH, Witte W.** Livestock-associated MRSA: the impact on humans. *Antibiotics (Basel)* 4, 521–543, 2015
- DairyNZ.** *New Zealand Dairy Statistics 2017-18.* DairyNZ Limited, Hamilton, New Zealand, 2017
- Danon L, Ford AP, House T, Jewell CP, Keeling MJ, Roberts GO, Ross JV, Vernon MC.** Networks and the epidemiology of infectious disease. *Interdiscip. Perspect. Infect. Dis.* 2011, e284909, 2011
- Devriese L, R Van Damme L, Fameree L.** Methicillin (Cloxacillin)-Resistant *Staphylococcus aureus* strains Isolated from Bovine Mastitis Cases. *J. Vet. Med.. Ser. B* 19, 598–605, 1972
- Eberhart RJ.** management of dry cows to reduce mastitis. *J. Dairy Sci.* 69, 1721–1732, 1986

- Fluit AC. Livestock-associated *Staphylococcus aureus*. *Clin. Microbiol. Infect.* 18, 735–744, 2012
- Gonçalves da Silva A, Baines SL, Carter GP, Heffernan H, French NP, Ren X, Seemann T, Bulach D, Kwong J, Stinear TP, Howden BP, Williamson DA. A phylogenomic framework for assessing the global emergence and evolution of clonal complex 398 methicillin-resistant *Staphylococcus aureus*. *Microb. Genom.* 3, e000105-22, 2017
- Gordon NC, Price JR, Cole K, Everitt R, Morgan M, Finney J, Kearns AM, Pichon B, Young B, Wilson DJ, Llewelyn MJ, Paul J, Peto TEA, Crook DW, Walker AS, Golubchik T. Prediction of *Staphylococcus aureus* antimicrobial resistance by whole-genome sequencing. *J. Clin. Microbiol.* 52, 1182–1191, 2014
- Grinberg A, Biggs PJ, Zhang J, Ritchie S, Oneroa Z, O'Neill C, Karkaba A, Velathanthiri NS, Coombs GW. Genomic epidemiology of methicillin-susceptible *Staphylococcus aureus* across colonisation and skin and soft tissue infection. *J. Infect.* 75, 326–335, 2017
- Grinberg A, Hittman A, Leyland M, Rogers L, Quesne BL. Epidemiological and molecular evidence of a monophyletic infection with *Staphylococcus aureus* causing a purulent dermatitis in a dairy farmer and multiple cases of mastitis in his cows. *Epidemiol. Infect.* 132, 507–513, 2004
- Grinberg A, Kingsbury DD, Gibson IR, Kirby BM, Mack HJ, Morrison D. Clinically overt infections with methicillin-resistant *Staphylococcus aureus* in animals in New Zealand: a pilot study. *N. Z. Vet. J.* 56, 237–242, 2008
- Guinane CM, Ben Zakour NL, Tormo-Mas MA, Weinert LA, Lowder BV, Cartwright RA, Smyth DS, Smyth CJ, Lindsay JA, Gould KA, Witney A, Hinds J, Bollback JP, Rambaut A, Penadés JR, Fitzgerald JR. Evolutionary genomics of *Staphylococcus aureus* reveals insights into the origin and molecular basis of ruminant host adaptation. *Genome Biol. Evol.* 2, 454–466, 2010
- Hata E, Katsuda K, Kobayashi H, Uchida I, Tanaka K, Eguchi M. Genetic variation among *Staphylococcus aureus* strains from bovine milk and their relevance to methicillin-resistant isolates from humans. *J. Clin. Microbiol.* 48, 2130–2139, 2010

- Heffernan H, Bakker S, Woodhouse R, Dyet K, Williamson DA. *Demographics, antimicrobial susceptibility and molecular epidemiology of Staphylococcus aureus in New Zealand, 2014*. Institute of Environmental Science and Research Ltd. Porirua, New Zealand, 2015.
- Hennekinne J-A, De Buyser M-L, Dragacci S. *Staphylococcus aureus* and its food poisoning toxins: characterization and outbreak investigation. *FEMS Microbiol. Rev.* 36, 815–836, 2012
- Hillerton E, Allison A. *Antibiotic resistance: challenges and opportunities*. New Zealand Veterinary Association. Price Waterhouse Coopers (PwC), New Zealand, 2015.
- Jarynowski A, Liljeros F. *Contact networks and the spread of MRSA in Stockholm hospitals*. Proceedings at the Second European Network Intelligence Conference, Institute of Electrical and Electronics Engineers, Pp 150–154, Karlskrona, Sweden, 2015
- Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S, Courtot M, Lo R, Williams LE, Frye JG, Elsayegh T, Sardar D, Westman EL, Pawlowski AC, Johnson TA, Brinkman FSL, Wright GD, McArthur AG. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, 566–573, 2017
- de Knecht LV, Kudirkiene E, Rattenborg E, Sørensen G, Denwood MJ, Olsen JE, Nielsen LR. Combining *Salmonella dublin* genome information and contact-tracing to substantiate a new approach for improved detection of infectious transmission routes in cattle populations. *Prev. Vet. Med.* 101, 73–83, 2018
- Kossinets G. Effects of missing data in social networks. *Social Networks* 28, 247–268, 2006
- Kruskal JB, Wish M. *Multidimensional Scaling*. Sage Publications, Newbury Park, California, US, 1978.
- Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874, 2016
- Kwak YG, Truong-Bolduc QC, Bin Kim H, Song K-H, Kim ES, Hooper DC. Association of norB overexpression and fluoroquinolone resistance in clinical isolates of *Staphylococcus aureus* from Korea. *J. Antimicrob. Chemother.* 68, 2766–2772, 2013

- Lacy-Hulbert J, Blackwell M, McDougall S.** SmartSAMB: *The smart approach to minimising mastitis*. Proceedings of the Society of Dairy Cattle Veterinarians of the NZVA Annual Conference, Wellington, New Zealand, 2011
- Letunic I, Bork P.** Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, 242-245, 2016
- Lowder BV, Guinane CM, Zakour NLB, Weinert LA, Conway-Morris A, Cartwright RA, Simpson AJ, Rambaut A, Nübel U, Fitzgerald JR.** Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus*. *PNAS* 106, 19545–19550, 2009
- Maddison WP.** Testing character correlation using pairwise comparisons on a phylogeny. *J. Theor. Biol.* 202, 195–204, 2000
- McCarthy AJ, Lindsay JA.** The distribution of plasmids that carry virulence and resistance genes in *Staphylococcus aureus* is lineage associated. *BMC Microbiol.* 12, 104, 2012
- McDougall S.** Bovine mastitis: epidemiology, treatment and control. *N. Z. Vet. J.* 50, 81–84, 2002
- McDougall S, Arthur DG, Bryan MA, Vermunt JJ, Weir AM.** Clinical and bacteriological response to treatment of clinical mastitis with one of three intra-mammary antibiotics. *N. Z. Vet. J.* 55, 161–170, 2007
- McDougall S, Hussein H, Petrovski K.** Antimicrobial resistance in *Staphylococcus aureus*, *Streptococcus uberis* and *Streptococcus dysgalactiae* from dairy cows with mastitis. *N. Z. Vet. J.* 62, 68–76, 2014
- McMillan K, Moore SC, McAuley CM, Fegan N, Fox EM.** Characterization of *Staphylococcus aureus* isolates from raw milk sources in Victoria, Australia. *BMC Microbiol.* 16, e12866, 2016
- Mehndiratta PL, Bhalla P.** Use of antibiotics in animal agriculture & emergence of methicillin-resistant *Staphylococcus aureus* (MRSA) clones: Need to assess the impact on public health. *Indian J. Med. Res.* 140, 339–344, 2014
- Mohammed S, Nigatu S.** Review on livestock associated methicillin resistant *Staphylococcus aureus* and its zoonotic importance. *Inter. J. Microbiol. Res.* 6, 164-174, 2015

- Nadon C, Van Walle I, Gerner-Smidt P, Campos J, Chinen I, Concepcion-Acevedo J, Gilpin B, Smith AM, Kam KM, Perez E, Trees E, Kubota K, Takkinen J, Nielsen EM, Carleton H. PulseNet International: vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. *Euro Surveill.* 22, e30544, 2017
- Nazari R, Godarzi H, Rahimi Baghi F, Moeinrad M. Enterotoxin gene profiles among *Staphylococcus aureus* isolated from raw milk. *Iran J. Vet. Res.* 15, 409–412, 2014
- Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* 12, 443–451, 2011
- Obadia T, Opatowski L, Temime L, Herrmann J-L, Fleury E, Boëlle P-Y, Guillemot D. Interindividual contacts and carriage of methicillin-resistant *Staphylococcus aureus*: a nested case-control study. *Infect. Control Hosp. Epidemiol.* 36, 922–929, 2015
- Ojo KK, Tung D, Luis H, Bernardo M, Leitao J, Roberts MC. Gram-positive *merA* gene in gram-negative oral and urine bacteria. *FEMS Microbiol. Lett.* 238, 411–416, 2004
- Olson ND, Lund SP, Colman RE, Foster JT, Sahl JW, Schupp JM, Keim P, Morrow JB, Salit ML, Zook JM. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. *Front. Genet.* 6, e00235, 2015
- Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693, 2015
- Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* 2, e000056, 2016
- Pei S, Morone F, Liljeros F, Makse H, Shaman JL. Inference and control of the nosocomial transmission of methicillin-resistant *Staphylococcus aureus*. *eLife* 7, e40977, 2018
- Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526–528, 2018

- Petrovski KR, Grinberg A, Williamson NB, Abdalla ME, Lopez-Villalobos N, Parkinson TJ, Tucker IG, Rapnicki P. Susceptibility to antimicrobials of mastitis-causing *Staphylococcus aureus*, *Streptococcus uberis* and *Str. dysgalactiae* from New Zealand and the USA as assessed by the disk diffusion test. *Aust. Vet. J.* 93, 227–233, 2015
- Petrovski KR, Heuer C, Parkinson TJ, Williamson NB. The incidence and aetiology of clinical bovine mastitis on 14 farms in Northland, New Zealand. *N. Z. Vet. J.* 57, 109–115, 2009
- Petrovski KR, Laven RA, Lopez-Villalobos N. A descriptive analysis of the antimicrobial susceptibility of mastitis-causing bacteria isolated from samples submitted to commercial diagnostic laboratories in New Zealand (2003-2006). *N. Z. Vet. J.* 59, 59–66, 2011
- Petrovski KR, Trajcev M, Buneski G. A review of the factors affecting the costs of bovine mastitis. *J. S. Afr. Vet. Assoc.* 77, 52–60, 2006
- Price LB, Stegger M, Hasman H, Aziz M, Larsen J, Andersen PS, Pearson T, Waters AE, Foster JT, Schupp J, Gillece J, Driebe E, Liu CM, Springer B, Zdovc I, Battisti A, Franco A, Źmudzki J, Schwarz S, Butaye P, Jouy E, Pomba C, Porrero MC, Ruimy R, Smith TC, Robinson DA, Weese JS, Arriola CS, Yu F, Laurent F, Keim P, Skov R, Aarestrup FM. *Staphylococcus aureus* CC398: host adaptation and emergence of methicillin resistance in livestock. *mBio* 3, e00305-11, 2012
- R-Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2018. URL: <https://www.R-project.org/>
- Rall VLM, Miranda ES, Castilho IG, Camargo CH, Langoni H, Guimarães FF, Araújo Júnior JP, Fernandes Júnior A. Diversity of *Staphylococcus* species and prevalence of enterotoxin genes isolated from milk of healthy cows and cows with subclinical mastitis. *J. Dairy Sci.* 97, 829–837, 2014
- Riva A, Borghi E, Cirasola D, Colmegna S, Borgo F, Amato E, Pontello MM, Morace G. Methicillin-resistant *Staphylococcus aureus* in raw milk: prevalence, SCCmec typing, enterotoxin characterization, and antimicrobial resistance patterns. *J. Food Prot.* 78, 1142–1146, 2015

- Rossi BF, Bonsaglia ECR, Castilho IG, Dantas STA, Salina A, Langoni H, Pantoja JCF, Budri PE, Fitzgerald-Hughes D, Júnior AF, Rall VLM. Genotyping of long-term persistent *Staphylococcus aureus* in bovine subclinical mastitis. *Microb. Pathog.* 132, 45–50, 2019
- Sakwinska O, Giddey M, Moreillon M, Morisset D, Waldvogel A, Moreillon P. *Staphylococcus aureus* host range and human-bovine host shift. *Appl. Environ. Microbiol.* 77, 5908–5915, 2011
- Saunders NA, Holmes A. Multilocus Sequence Typing (MLST) of *Staphylococcus aureus*. *Methods Mol. Biol.* 391, 71–85, 2007
- Seegers H, Fourichon C, Beaudeau F. Production effects related to mastitis and mastitis economics in dairy cattle herds. *Vet. Res.* 34, 475–491, 2003
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069, 2014
- Sharma M, Nunez-Garcia J, Kearns AM, Doumith M, Butaye PR, Argudín MA, Lahuerta-Marin A, Pichon B, AbuOun M, Rogers J, Ellis RJ, Teale C, Anjum MF. Livestock-Associated Methicillin Resistant *Staphylococcus aureus* (LA-MRSA) Clonal Complex (CC) 398 Isolated from UK Animals belong to European Lineages. *Front. Microbiol.* e01741, 2016
- Souvorov A, Agarwala R, Lipman DJ. SKESA: strategic *k-mer* extension for scrupulous assemblies. *Genome Biol.* 19, 153–166, 2018
- Spoor LE, McAdam PR, Weinert LA, Rambaut A, Hasman H, Aarestrup FM, Kearns AM, Larsen AR, Skov RL, Fitzgerald JR. Livestock origin for a human pandemic clone of community-associated methicillin-resistant *Staphylococcus aureus*. *mBio* 4, e00356, 2013
- Stapleton P, Pike R, Mullany P, Lucas V, Roberts G, Rowbury R, Wilson M, Richards H. Mercuric resistance genes in gram-positive oral bacteria. *FEMS Microbiol. Lett.* 236, 213–220, 2004
- Steele N, McDougall S. Effect of prolonged duration therapy of subclinical mastitis in lactating dairy cows using penethamate hydriodide, *N. Z. Vet. J.* 62, 38–46, 2014
- Sung JM-L, Lloyd DH, Lindsay JA. *Staphylococcus aureus* host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. *Microbiol.* 154, 1949–1959, 2008

- van den Borne BHP, Nielen M, van Schaik G, Melchior MB, Lam TJGM, Zadoks RN. Host adaptation of bovine *Staphylococcus aureus* seems associated with bacteriological cure after lactational antimicrobial treatment. *J. Dairy Sci.* 93, 2550–2558, 2010
- Verkade E, van Benthem B, den Bergh MK, van Cleef B, van Rijen M, Bosch T, Kluytmans J. Dynamics and determinants of *Staphylococcus aureus* carriage in livestock veterinarians: a prospective cohort study. *Clin. Infect. Dis.* 57, 11–17, 2013
- Wasserman S, Faust K. *Social network analysis: methods and applications*. Cambridge University Press, Cambridge, UK, 1994
- Watts JL. Etiological agents of bovine mastitis. *Vet. Microbiol.* 16, 41–66, 1988
- Williamson DA, Roberts SA, Ritchie SR, Coombs GW, Fraser JD, Heffernan H. Clinical and molecular epidemiology of methicillin-resistant *Staphylococcus aureus* in New Zealand: rapid emergence of Sequence Type 5 (ST5)-SCCmec-IV as the dominant community-associated MRSA Clone. *PLoS One* 8, e62020-27, 2013
- Wilson DJ. Insights from genomics into bacterial pathogen populations. *PLoS Pathog.* 8, e1002874-83, 2012
- World Health Organisation (WHO). *WHO critically important antimicrobials for human medicine 6th revision*. Available at: <https://www.who.int/foodsafety/publications/antimicrobials-sixth/en/> Accessed 2 October 2019. World Health Organization, Utrecht, Netherlands, 2018
- Wulf M, Voss A. MRSA in livestock animals-an epidemic waiting to happen? *Clin. Microbiol. Infect.* 14, 519–521, 2008
- Yan B, Gregory S. Finding missing edges and communities in incomplete networks. *J. Phys. A: Math. Theor.* 44, e495102, 2011
- Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644, 2012
- Zhang J, Xiong Y, Rogers L, Carter GP, French N. Genome-by-genome approach for fast bacterial genealogical relationship evaluation. *Bioinformatics* 34, 3025–3027, 2018

General Discussion

8.1. Introduction

The ability to infer the transmission dynamics of a disease outbreak is a key aspect in many epidemiological investigations; however, outbreak reconstruction is often time-consuming, expensive and results in many uncertainties. Contact networks are a popular epidemiological tool that are often used to inform the population contact structure in disease simulation models; supporting inferences on who-infected-whom and providing evidence for or against different disease control strategies (Welch *et al.* 2011; Ray *et al.* 2016). Nevertheless, using contact networks can also have many limitations, and in the real-world network data is often incomplete or unreliable (Craft, 2015) making it difficult to infer the transmission dynamics from network data alone. Therefore, a growing number of studies are developing methods to use the molecular sequence data of rapidly evolving pathogens in outbreak reconstruction (Campbell *et al.* 2018), however despite the success of previous studies, using molecular sequence data also presents many challenges.

For example, many of the current methods often rely on a high proportion of disease cases having been identified and sampled, as well as detailed knowledge surrounding epidemiological parameters such as the generation time (*i.e.*, the time between an individual becoming infected and infecting others). Incorporating molecular data also often requires both an evolutionary model, indicating how and at what rate mutations occur, and a pathogen population model to specify the transmission dynamics within and between host populations such as the number of lineages present, and how lineages are transmitted upon infection. For these reasons, it is important to develop methods that integrate pathogen molecular sequence data with epidemiological data, including contact networks, and clinical information to obtain more robust inferences on the origin and transmission of a disease, whilst also providing insight into the evolution of important pathogen traits such as antimicrobial resistance (Wilson, 2012; Kao *et al.* 2014; Campbell *et al.* 2019).

This general discussion brings together the context of this thesis, which aimed at investigating the transmission and evolution of two important pathogens; *Campylobacter jejuni* and *Staphylococcus aureus*, through the integration of pathogen whole genome sequence (WGS) data and contact network data. In this discussion the contents of each chapter have been critically evaluated with demonstrations of how the results fit within the current knowledge and support potential areas for future research.

8.2. Overview of results: *Campylobacter jejuni*

In 2014, a previously unidentified *C. jejuni* lineage, known as ST-6964, was detected across all major poultry suppliers in New Zealand, and has since been associated with a number of human campylobacteriosis sporadic cases and outbreaks. In addition to the rapid spread of this lineage, antimicrobial drug-resistance was also observed in many of the disease cases, and subsequent genomic analysis revealed a number of evolutionary changes associated with tetracycline and fluoroquinolone resistance (French *et al.* 2019). Despite this knowledge, many questions still remain about the means by which this lineage was able to spread between the different poultry suppliers, when previous dominant lineages, such as ST-474 and ST-48, were shown to be strongly associated with single suppliers (Müllner *et al.* 2010). This change in epidemiology suggests there could be a potential gap in biosecurity providing a pathway for *C. jejuni* to spread between farms belonging to different poultry suppliers.

In Chapter 3, the results from an industry survey suggest that the biosecurity practices between different producers vary greatly, with many farms forgoing basic recommended practices, and although this low level of biosecurity does not provide any evidence for the spread of *C. jejuni*, it is clear that more needs to be done to promote good practices across the industry and maintain a minimum level of biosecurity in the event of a disease outbreak. Nevertheless, current evidence suggests that improving the general standard of biosecurity may not be enough to reduce the public health risk of *Campylobacter*, and

instead a more targeted approach is needed. For example, a study by Ridley and colleagues (2011) compared the *Campylobacter* prevalence between farms with standard biosecurity practices versus farms with enhanced biosecurity practices, and although there was a reduction in *Campylobacter* prevalence on the majority of farms with enhanced biosecurity measures, in many cases it was not enough to prevent colonization (Ridley *et al.* 2011). A similar pattern can be seen in many countries where, despite the introduction of evidence-based interventions throughout their poultry industry together with increased media coverage educating the consumer on the risk of *Campylobacter*, human campylobacteriosis cases remain high. This emphasises the difficulties in controlling *Campylobacter* and the importance of identifying potential transmission pathways that could be targeted.

In addition to the disease risk pathways that are commonly associated with production processes, additional risk pathways have been observed due to the interaction between commercial poultry and both backyard poultry (Conan *et al.* 2012) and wild birds (Gilchrist, 2007; Si *et al.* 2013), although these pathways can be notoriously more difficult to characterise. In Chapter 4, results provide evidence of a spatial overlap between backyard poultry with both commercial poultry and wild birds; a contact pattern that presents a high risk for the transmission of endemic diseases as well as the introduction of exotic diseases to the commercial poultry industry. These results are similar to that of other studies presenting the growing risk of backyard poultry as disease reservoirs for not only avian diseases that pose a risk to the commercial poultry industry but also many zoonotic diseases that may have huge public health consequences (Behravesh *et al.* 2014; Pohjola *et al.* 2016; Derksen *et al.* 2018). Long-range movements further demonstrate the importance of considering backyard poultry when planning disease surveillance and control activities even after taking into account limitations in the data, as previously discussed, as well as highlighting the potential of using online trading platforms to start educating and promoting good practices for animal health and biosecurity.

To further characterise the risk pathways in the New Zealand commercial poultry industry, Chapter 5, uses a distance-based linear model to determine the contribution of different network and geographical distances to the pairwise genetic distance between 167 ST-6964 *C. jejuni* isolates. Results suggested that both transporting feed vehicles and local transmission mechanisms may have played an important role in the spread of this pathogen between farms and suppliers. Both these risk factors have been previously identified and many studies emphasise the importance of expanding biosecurity measures beyond the production area to include the disinfection of important contact points such as farm gates and vehicles (Newell *et al.* 2011; Silbanda *et al.* 2018). However, the cost-effectiveness of these measures has also been questioned (Hald *et al.* 2000), and before making recommendation to producers it would be important to complete additional environmental sampling; such as those from vehicle tyres or personnel, to provide more robust evidence of these risk factors. Additional sampling may also help to discern between the different local mechanisms, for instance, a study by Ridley and colleagues (2011) was able to match *Campylobacter* sequences sampled from farms with those isolated from the lunch bags of employees; emphasizing the movement of personnel as a major risk factor (Ridley *et al.* 2011). Other studies have managed to make a distinction between employees, highlighting depopulation and the entrance of the catching crew as a major risk for the introduction of *Campylobacter*, leading to the conclusion that the most cost-effective intervention might be the introduction of a strict all-in-all-out system with a complete ban on partial depopulation (Ellis-Iversen *et al.* 2012; van Wagenberg *et al.* 2016). In New Zealand, this would require a major restructure for many poultry producers who currently rely on mixed-aged production cycles or practice partial depopulation in order to optimise production and meet the requirement for birds of particular sizes by retailers and fast food companies. Therefore, before any recommendation can be made, it would be important for future studies to collect additional isolates for sequencing to help determine the importance between different local pathways.

Together with more *C. jejuni* isolates, it would also be important for future studies to collect additional farm demographic and contact data; as demonstrated in Chapter 6, in which the inferences about the spread of *C. jejuni* are limited by the available data. Ideally, any study aiming to use pathogen sequence data in an epidemiological analysis should design the study to ensure that such population level data is collected in parallel to the isolates intended for sequencing thereby limiting the mismatch between datasets (*i.e.*, not knowing the contact structure for the sampled pathogen, or lacking an appropriate pathogen for an observed network) (Craft, 2015; Eames *et al.* 2015). However, it also remains clear that there is a need for the further development of robust analytical methods that aim at combining network-based disease transmission models into a phylogenetic framework in order to make accurate inference about both pathogen transmission and evolutionary dynamics.

8.3. Overview of results: *Staphylococcus aureus*

Bovine mastitis causes both major economic losses and animal welfare concerns for dairy industries worldwide (Petrovski *et al.* 2006; Halasa *et al.* 2007), with *S. aureus* known to be one of the primary pathogens responsible for both clinical and sub-clinical mastitis on many dairy herds. In New Zealand, an estimated quarter of all clinical mastitis cases are thought to be caused by an infection with *S. aureus* (Petrovski *et al.* 2009; Notcovich *et al.* 2018). Until recently, these infections would be treated with the use of blanket antimicrobial dry-cow therapy and the treatment of intra-mammary infection during drying off. However, with growing concerns for the use of antimicrobials in food-producing animals, specifically those critical to human medicine, new recommendations have been developed to promote the stewardship of antimicrobials, and support the goal to eliminate the use of antimicrobials for the maintenance of animal health and wellness across New Zealand by 2030 (Hillerton and Allison, 2015). Therefore, to effectively control mastitis whilst also ensuring the stewardship of antimicrobials, it is crucial to have a good understanding of both the transmission dynamics of the pathogens responsible

including knowledge on the pathways that may be contributing to their spread and maintenance within the population, and the current evolutionary mechanisms that present a potential risk of antimicrobial resistance.

In chapter 7, the genetic relatedness between 59 *S. aureus* isolates sampled from 17 dairy farms located in the Waikato region of New Zealand was determined in order to investigate both the role of live animal movements towards the local spread of *S. aureus* and contribute to findings on the antimicrobial resistance patterns within the pathogen population. Overall, very little diversity was found between the isolates, although only a limited number of the study farms could be directly connected via an animal movement. This result suggests that other local transmission mechanisms, such as the movement of personnel between farms, could be more important for the spread of *S. aureus*; particularly within high risk groups such as casual employees. Many previous studies have provided evidence of transmission between humans and animals (Juhász-Kaszanyitzky *et al.* 2007; Türkyılmaz *et al.* 2010); however, the risk of human movement between farms as a transmission pathway has not been fully characterised and without further sampling across a range of different environments, hosts and other potential fomites, no further inferences can be made about the contribution of different risk pathways. Nevertheless, it is clear that using animal movement data alone may not be enough to fully capture the transmission dynamics of *S. aureus*. Future research should focus on potentially high-risk groups for example, in New Zealand there are many Willing Workers on Organic Farms (*i.e.*, ‘WWOOFers’), that exchange hours of work for accommodation and food. However, despite many WWOOFers moving between a number of different farms over long distances (<https://wwoofinternational.org/>), no research has been conducted to look at their potential impact on the spread of disease. This knowledge gap may exist for many reasons, with a number of limitations on our ability to capture contacts between humans and animals. This includes contact between humans and livestock, pets and wildlife, that could present a potential risk for both

zoonotic and reverse-zoonotic transmission of *S. aureus* (Gonçalves *et al.* 2017); including the potential transmission of antimicrobial resistance (Cuny *et al.* 2015).

For this reason, Chapter 7 further explores the genetic relatedness between the 59 bovine derived *S. aureus* isolates with both national and international *S. aureus* isolates previously collected from human, pets and livestock. Results from this analysis show the majority of New Zealand bovine isolates within the same genetic cluster, supporting previous evidence for the limited transmission of *S. aureus* between humans and cattle (Burgess and French, 2017). However, the genetic relatedness between the New Zealand bovine isolates and a single bovine isolate originating from Australia does suggest a potential transmission event between the two countries, although with only a limited amount of data, no inferences can be made as to the directionality, timing, or mechanism of spread. Nevertheless, this result does highlight the importance of maintaining good animal movement records for not only in the event of an outbreak but also to study historic transmission events that may be responsible for the long-term maintenance of a disease within a population.

However, despite this importance, many databases aimed at recording movement data are known to only capture a proportion of all the animal movements; making it difficult to reliably infer disease dynamics. For example, in New Zealand the recent outbreak of *Mycoplasma bovis* brought to attention limitations within the National Animal Identification and Tracing (NAIT) system that was used in the outbreak response including non-compliance among producers, mismatches between information recorded in different systems *i.e.*, NAIT versus LIC MINDA data, and incorrect or missing records (*i.e.*, recorded culls for animals still alive and being traded); making it difficult to rely solely on NAIT data to have captured all movements relevant for transmission, and therefore increasing the time and resources needed to trace all possible contacts (Browning *et al.* 2019).

8.4. Future opportunities

Many of the findings presented in this thesis have highlighted the potential of using both network data and pathogen whole genome sequence data to investigate pathogen transmission and evolutionary dynamics; adding not only to areas of knowledge within epidemiology, phylogenetics and public health, but also highlighting opportunities for future research within these fields. Across Chapters 3 to 6 the structure and vulnerability of New Zealand's poultry industry to disease incursion was demonstrated; with results from an industry survey conducted in 2016 highlighting the varying level of biosecurity among poultry producers. Without further research, it is hard to understand what drives a producer to adopt one biosecurity measure over another, although thematic analysis on the qualitative free-text survey responses show that producers would like to see more evidence-based research into the effectiveness of different biosecurity measures. This suggests that some producers may currently have doubts that prevent them from utilising a full range of biosecurity practices. The generally low level of biosecurity across the industry will also become of greater concern as more producers transition to free-range housing systems in order to meet the animal welfare legislations by 2022 (MPI, 2017), making it more difficult to maintain adequate biosecurity due to a greater risk of disease incursions particularly for diseases such as avian influenza that can be easily spread from wild birds (Burns *et al.* 2013). Given these concerns, a new two-year project in collaboration between Massey University's EpiCentre, the Poultry Industry Association of New Zealand (PIANZ), and the Egg Producers Federation (EPF) aims to use semi-structured interviews of both commercial and backyard poultry producers to co-design a national poultry data system. The aim of this database would be to not only capture important data on farm demographics, contact patterns and biosecurity practices that could be used in response to an outbreak, but also to support producers in maintaining good biosecurity practices and increase future engagement between producers and researchers by introducing feedback loops that clearly demonstrate the value of producer responses; with aims of encouraging their future participation in research.

In addition to these targets, a key aim of the project is to not only foster communication between researchers, industry, and commercial producers but also backyard poultry traders. In Chapter 4, the use of data from the auction website TradeMe®, showed a highly active backyard poultry industry in New Zealand with more than 13,000 traders and 16,000 trades occurring each year through TradeMe® alone. These trades present a significant risk for both the spread of endemic diseases as well as the introduction of exotic diseases, due to the spatial overlap of many wild bird populations and backyard poultry producers. Furthermore, backyard poultry producers are notoriously more difficult to engage with since there is no national industry representative body and no legislative requirement for owners to register their poultry unless they keep more than 20 birds on site; making it difficult to fully capture movements within the backyard poultry network and characterise the risk they present (Burns *et al.* 2013). For these reasons, it is clear that further engagement with backyard poultry producers is needed across different platforms, such as TradeMe®, in order to support disease control and surveillance activities including disease simulation modelling, contact tracing and, resource allocation.

Alongside additional network data, the continued use of WGS data for endemic pathogens such as *Campylobacter* would be of great value, as demonstrated in Chapter 5 which integrated both pathogen sequence data and network data into a distance-based linear model to investigate the role of different transmission pathways on the spread of a recently emerged *C. jejuni* lineage. After controlling for parent company, a significant association was found between the pairwise genetic distance between isolates and both the road distance and the network distance of transporting feed vehicles. These findings provide insight into potential pathways where control measures may currently be non-existent or insufficient in mitigating transmission. For example, the association with road distance may be capturing the movement of personnel that present a risk for between farm transmission. However, before any further inferences can be made it would be

essential to increase the number of sequenced isolates, improving phylogenetic reconstruction, and encourage greater compliance from producers in order to eliminate the need to infer or impute network features. Only with more data can any insight into the spatial and temporal transmission patterns be extended to make any inference on who-infected-whom (Astill *et al.* 2018); further highlighting the importance of continuing surveillance in not only commercial poultry, but also a need to introduce surveillance in backyard poultry and wild bird populations, particularly those that are either considered high-risk such as water birds or those captured in commercial production areas.

The need for further surveillance and the continued development of databases that can capture host contact patterns is also highlighted in Chapter 7 in which multiple phylogenetic analyses across different geographical scales were used to investigate the transmission dynamics of *S. aureus*. Results showed very little genetic diversity between locally sourced bovine-derived isolates however, without a reliable contact network it was difficult to infer transmission dynamics. A greater number of sequenced isolates targeting those farms known to be connected in the network would help to identify the contribution of live animal movements to the spread of *S. aureus*. Further phenotypic sensitivity tests such as zone diffusion or minimum inhibitory concentrations (MIC) could also be performed to help determine the relationship between the isolates, as well as provide sensitivity data for the different antimicrobials; an important step in determining the resistance patterns in a population that may be used to guide local treatment and ensure the stewardship of antimicrobials. Sensitivity data could also be used to make a comparison among the global isolates as both in order to help determine where isolates originate from (Nishi *et al.* 2016), and support evidence for the transportation of *S. aureus* between geographical locations as suggested by the number of human isolates sampled from different countries that cluster in the phylogeny. MIC values have also been used to compare resistance profiles at geographical locations and

antimicrobial usage patterns (de Oliveira *et al.* 2000), although these studies often involve extensive sampling with many aiming to limit isolates from the same herd to improve the comparison between different countries (Aarestrup *et al.* 2012; Stefani *et al.* 2012) alongside collecting additional data in order to identify individual management practices that can influence antimicrobial resistance levels (Østerås *et al.* 1999; de Oliveira *et al.* 2000). This study design increases both the cost and time needed to collect samples which may be hard to justify in New Zealand where there is very little evidence of MRSA (Petrovski *et al.* 2011; Williamson *et al.* 2013) and the presence of other resistance genes remains low. Nevertheless, the increasing global prevalence of LA-MRSA (Mehndiratta and Bhalla, 2014; Cuny *et al.* 2015; Mohammed and Nigatu, 2015) as well as evidence of resistance in gut bacteria (Toombs-Ruane *et al.* 2017) highlights the importance of antimicrobial stewardship, limiting the exposure of *S. aureus* infected animals to antimicrobial drugs, and the widespread adoption of mastitis control programs to prevent infections.

8.5. Data limitations

Despite the differences between the two pathogens examined in this thesis, there are many commonalities in the limitations of the datasets that should be considered or addressed in future studies. For example, both the poultry and cattle contact networks were incomplete although for different reasons. For the poultry network, the low response rate to the industry survey contributed substantially to missing links in the network although most of the nodes (*i.e.*, producers) were correctly identified through the PIANZ database. On the other hand, for the dairy cattle network, it is a challenge being able to identify all the dairy farms in New Zealand and there is known underreporting of certain types of cattle movements resulting a contact network that may be incomplete or non-representative. Both networks were also impacted by the lack of knowledge about the range of alternative transmission pathways for the disease such as the the contribution of different wildlife reservoirs (*i.e.*, wild birds versus rodents versus insects) for

Campylobacter that make it difficult not only to fully characterise the network but also use a modelling approach to study disease transmission dynamics. Even in cases where risk pathways are well characterised, additional information may be needed to include them in transmission models. For example, it is well recognised that catchers entering poultry sheds are associated with the introduction of *Campylobacter* in a flock, however, more information is needed on the catching protocol on each farm to model this risk further such as if the farm uses a catching company and when are the first cuts taken from each shed. Similar information regarding the movements of personnel on dairy farms would also be highly valuable to inform disease models particularly if employees have direct contact with animals and frequently move between farms if they are working for different businesses on casual employment contracts.

In addition to the limitations in constructing the contact networks, there were also issues arising with the phylogenetic analysis due to the restricted number and spatial distribution of the pathogen isolate samples. For many studies, this is often the case due to time and cost restraints, and although comparisons can be made with samples from across multiple studies, such as with *S. aureus*, it is often difficult to determine the epidemiological linkages across the different study samples. Therefore even in the presence of a phylogeny, it is important to ask if the isolates are relevant for disease transmission. For some diseases this may be more difficult if the transmission pathways are not well described, however in most cases, this limitation can be overcome by strategically sampling farms with known temporal movement connections to see if the isolates are more closely related in comparison to farms with no known movement connections. For instance, in this thesis *S. aureus* was selected as a pathogen example as its transmission dynamics have been well described with most transmission events occurring in the milking shed and between farm transmission predominantly occurring through the movement of adult lactating dairy cattle (Álvarez *et al.* 2011), which is generally well captured in MINDA. This means that the contact network relevant to *S.*

aureus transmission should be relatively complete, however, the main limitation with the phylogenetic data in this case was that few of the sampled herds had any direct movement connections with each other. This emphasises the importance of study design and sampling strategy when trying to integrate contact networks and pathogen phylogenies.

8.6 Developing methodologies

However, even after giving consideration for the study design many studies focusing on livestock diseases still lack a “true” dataset since most of the time there is always missing information on infected individuals, their contacts, and who they infected. Without a “true” dataset, it is difficult to validate new methods and test their assumptions, although in more recent years, the emergence of novel diseases such as COVID-19 (Andersen *et al.* 2020) has given rise to an influx of contact data. These improvements in contact tracing alongside the growing use of online repositories that contain information on genomic variants, present new opportunities to validate current methods. However, with such data becoming increasingly accessible, the importance of maintaining transparency and reproducibility in research grows too (Catalá-López *et al.* 2016; Prager *et al.* 2019). For example, many of the findings presented in this thesis have highlighted the potential of using both network data and pathogen whole genome sequence data to investigate the transmission and evolutionary dynamics of two specific pathogens and the methods could easily be extended to other pathogens. However, before this is done, it is important that researchers consider the limitation in their own data, and if the assumptions made in each method are appropriate for the study design. For example, the *phybreak* model, used in Chapter 6, assumes the Jukes-Cantor nucleotide substitution model under which only a single mutation rate is inferred (Klinkenberg *et al.* 2017). Therefore, it is important that if this method is used, there is an understanding of the differences between different substitution models, which is most relevant to the pathogen being studied, and how this assumption may limit the results.

8.7. Concluding comments

In the world of big data there is ever increasing opportunities to apply multidisciplinary techniques across a range of data sources. A good example can be seen in the number of methodologies recently developed integrating both pathogen sequence data and network analysis; representing powerful new tools in the fight against infectious disease. However, many of these approaches have yet to be validated, and with many relying on a number of simplifying assumptions it is clear that further discussion is needed to ensure the development of not only robust methods but also reliable databases that are often of interest across multiple research disciplines.

8.8. References

- Aarestrup FM, Brown EW, Detter C, Gerner-Smidt P, Gilmour MW, Harmsen D, Hendriksen RS, Hewson R, Heymann DL, Johansson K, Ijaz K, Keim PS, Koopmans M, Kroneman A, Lo Fo Wong D, Lund O, Palm D, Sawanpanyalert P, Sobel J, Schlundt J. Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. *Emerg. Infect. Dis.* 18, e1, 2012
- Álvarez L, Webb C, Holmes M. A novel field-based approach to validate the use of network models for disease spread between dairy herds. *Epidemiol. Infect.*, 139, 1863-1874, 2011
- Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nat. Med.* 26, 450–452, 2020
- Astill J, Fraser E, Dara R, Sharif S. Detecting and Predicting Emerging Disease in Poultry with the Implementation of New Technologies and Big Data: A Focus on Avian Influenza Virus. *Front. Vet. Sci.* 5, 263-275, 2018
- Behravesh CB, Brinson D, Hopkins BA, Gomez TM. Backyard poultry flocks and salmonellosis: a recurring, yet preventable public health challenge. *Clin. Infect. Dis.* 58, 1432–1438, 2014
- Browning G, Caswell J, Cobb S, Fox L, French NP, Garner G, Humphris M, Madin B, McDougall S, Morton J, Nichols R. *Report of the Mycoplasma bovis technical advisory group (TAG) in response to the terms of reference (June 2019)*. Mycoplasma bovis technical advisory group, Wellington, New Zealand, 2019
- Bull SA, Allen VM, Domingue G, Jorgensen F, Frost JA, Ure R, Whyte R, Tinker D, Corry JEL, Gillard-King J, Humphrey TJ. Sources of *Campylobacter* spp. Colonizing Housed Broiler Flocks during Rearing. *Appl. Environ. Microbiol.* 72, 645–652, 2006
- Burgess S, French NP. *Antimicrobial resistant bacteria in dairy cattle: A review*. New Zealand Food Safety Science and Research Centre, Wellington, New Zealand, 2017
- Burns TE, Ribble C, McLaws M, Kelton D, Stephen C. Perspectives of an underrepresented stakeholder group, backyard flock owners, on poultry health and avian influenza control. *J. Risk Res.* 16, 510–518, 2013

- Campbell F, Cori A, Ferguson N, Jombart T.** Bayesian inference of transmission chains using timing of symptoms, pathogen genomes and contact data. *PLoS Comput. Biol.* 15, e1006930, 2019
- Campbell F, Strang C, Ferguson N, Cori A, Jombart T.** When are pathogen genome sequences informative of transmission events? *PLoS Pathog.* 14, e1006885, 2018
- Catalá-López F, Moher D, Tabarés-Seisdedos R.** Improving transparency of scientific reporting to increase value and reduce waste in mental health research. *Rev. Psiquiatr. Salud Ment.* 9, 1-3, 2016
- Conan A, Goutard FL, Sorn S, Vong S.** Biosecurity measures for backyard poultry in developing countries: a systematic review. *BMC Vet. Res.* 8, 240-250, 2012
- Craft ME.** Infectious disease transmission and contact networks in wildlife and livestock. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, e20140107, 2015
- Cuny C, Wieler L, Witte W.** Livestock-Associated MRSA: The Impact on Humans. *Antibiotics* 4, 521–543, 2015
- Derksen T, Lampron R, Hauck R, Pitesky M, Gallardo RA.** Biosecurity Assessment and Seroprevalence of Respiratory Diseases in Backyard Poultry Flocks Located Close to and Far from Commercial Premises. *Avian Dis.* 62, 1–5, 2018
- Eames K, Bansal S, Frost S, Riley S.** Six challenges in measuring contact networks for use in modelling. *Epidemics* 10, 72–77, 2015
- Ellis-Iversen J, Ridley A, Morris V, Sowa A, Harris J, Atterbury R, Sparks N, Allen V.** Persistent environmental reservoirs on farms as risk factors for *Campylobacter* in commercial poultry. *Epidemiol. Infect.* 140, 916–924, 2012
- French NP, Zhang J, Carter GP, Midwinter AC, Biggs PJ, Dyet K, Gilpin BJ, Ingle DJ, Mulqueen K, Rogers LE, Wilkinson DA, Greening SS, Mullner P, Williamson DA.** Rapid and concurrent emergence of fluoroquinolone and tetracycline-resistant *Campylobacter jejuni* in humans and poultry in New Zealand. *Emerg. Infect. Dis.* 25, e2226-2234, 2019
- Gilchrist P.** Involvement of free-flying wild birds in the spread of the viruses of avian influenza, Newcastle disease and infectious bursal disease from poultry products to commercial poultry. *World's Poultry Sci. J.* 61, 198–214, 2005

- Gonçalves da Silva A, Baines SL, Carter GP, Heffernan H, French NP, Ren X, Seemann T, Bulach D, Kwong J, Stinear TP, Howden BP, Williamson DA. A phylogenomic framework for assessing the global emergence and evolution of clonal complex 398 methicillin-resistant *Staphylococcus aureus*. *Microb. Genom.* 3, e000105, 2017
- Halasa T, Huijps K, Østerås O, Hogeveen H. Economic effects of bovine mastitis and mastitis management: A review. *Vet. Quart.* 29, 18–31, 2007
- Hald B, Wedderkopp A, Madsen M. Thermophilic *Campylobacter* spp. in Danish broiler production: a cross-sectional survey and a retrospective analysis of risk factors for occurrence in broiler flocks. *Avian Pathol.* 29, 123–131, 2000
- Hillerton E, Allison A. *Antibiotic resistance: challenges and opportunities*. New Zealand Veterinary Association. Pricewaterhouse Coopers (PwC), New Zealand, 2015
- Juhász-Kaszanyitzky E, Jánosi S, Somogyi P, Dán A, van der Graaf-van Bloois L, van Duijkeren E, Wagenaar JA. MRSA transmission between cows and humans. *Emerg. Infect. Dis.* 13, 630–632, 2007
- Kao RR, Haydon DT, Lycett SJ, Murcia PR. Supersize me: how whole-genome sequencing and big data are transforming epidemiology. *Trends Microbiol.* 22, 282–291, 2014
- Klinkenberg D, Backer JA, Didelot X, Colijn C, Wallinga J. Simultaneous inference of phylogenetic and transmission trees in infectious disease outbreaks. *PLoS Comput. Biol.* 13, e1005495, 2017
- Mehndiratta PL, Bhalla P. Use of antibiotics in animal agriculture & emergence of methicillin-resistant *Staphylococcus aureus* (MRSA) clones: Need to assess the impact on public health. *Indian J. Med. Res.* 140, 339–344, 2014
- Mohammed S, Nigatu S. Review on Livestock Associated Methicillin Resistant *Staphylococcus aureus* and its Zoonotic Importance. *Int. J. Microbiol. Res.* 6, 164–174, 2015
- Ministry for Primary Industries (MPI). *Animal Welfare Regulations 2017; Regulatory Impact Statement*. Ministry for Primary Industries, Wellington, New Zealand, 2017

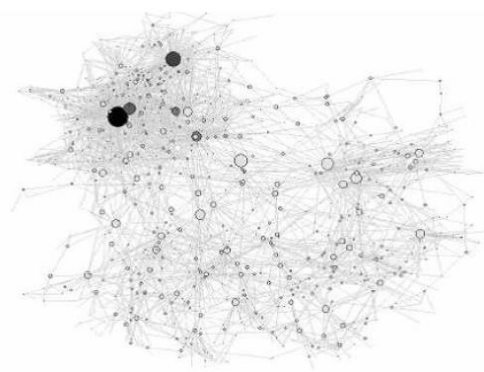
- Mullner P, Collins-Emerson JM, Midwinter AC, Carter P, Spencer SE, van der Logt P, Hathaway S, French NP.** Molecular epidemiology of *Campylobacter jejuni* in a geographically isolated country with a uniquely structured poultry industry. *Appl. Environ. Microbiol.* 76, 2145–54, 2010
- Newell DG, Elvers KT, Dopfer D, Hansson I, Jones P, James S, Gittins J, Stern NJ, Davies R, Connerton I, Pearson D, Salvat G, Allen VM.** Biosecurity-Based Interventions and Strategies to Reduce *Campylobacter* spp. on Poultry Farms. *Appl. Environ. Microbiol.* 77, 8605–8614, 2011
- Nishi A, Milner DA Jr, Giovannucci EL, Nishihara R, Tan AS, Kawachi I, Ogino S.** Integration of molecular pathology, epidemiology and social science for global precision medicine. *Expert Rev. Mol. Diag.* 16, 11–23, 2016
- Notcovich S, DeNicolo G, Flint SH, Williamson NB, Gedye K, Grinberg A, Lopez-Villalobos N.** Biofilm-Forming Potential of *Staphylococcus aureus* Isolated from Clinical Mastitis Cases in New Zealand. *Vet. Sci.* 5, 2018
- de Oliveira AP, Watts JL, Salmon SA, Aarestrup FM.** Antimicrobial Susceptibility of *Staphylococcus aureus* Isolated from Bovine Mastitis in Europe and the United States. *J. Dairy Sci.* 83, 855–862, 2000
- Østerås O, Edge VL, Martin SW.** Determinants of Success or Failure in the Elimination of Major Mastitis Pathogens in Selective Dry Cow Therapy. *J. Dairy Sci.* 82, 1221–1231, 1999
- Petrovski KR, Buneski G, Trajcev M.** A review of the factors affecting the costs of bovine mastitis: review article. *J. S. African Vet. Ass.* 77, 52–60, 2006
- Petrovski KR, Heuer C, Parkinson TJ, Williamson NB.** The incidence and aetiology of clinical bovine mastitis on 14 farms in Northland, New Zealand. *N. Z. Vet. J.* 57, 109–115, 2009
- Petrovski KR, Laven RA, Lopez-Villalobos N.** A descriptive analysis of the antimicrobial susceptibility of mastitis-causing bacteria isolated from samples submitted to commercial diagnostic laboratories in New Zealand (2003–2006). *N. Z. Vet. J.* 59, 59–66, 2011
- Pohjola L, Nykäsenoja S, Kivistö R, Soveri T, Huovilainen A, Hänninen ML, Fredriksson-Ahomaa M.** Zoonotic Public Health Hazards in Backyard Chickens. *Zoonoses Public Health* 63, 420–430, 2016

- Prager EM, Chambers KE, Plotkin JL, McArthur DL, Bandrowski AE, Martone ME, Bergstrom HC, Beshpalov A, Graf C.** Improving transparency and scientific rigor in academic publishing. *J. Neuro. Res.* 97, 377– 390, 2019
- Ray B, Ghedin E, Chunara R.** Network inference from multimodal data: A review of approaches from infectious disease transmission. *J. Biomed. Informatics* 64, 44–54, 2016
- Ridley A, Morris V, Gittins J, Cawthraw S, Harris J, Edge S, Allen V.** Potential sources of *Campylobacter* infection on chicken farms: contamination and control of broiler-harvesting equipment, vehicles and personnel. *J. Appl. Microbiol.* 111, 233–244, 2011
- Si Y, de Boer WF, Gong P.** Different Environmental Drivers of Highly Pathogenic Avian Influenza H5N1 Outbreaks in Poultry and Wild Birds. *PLoS One* 8, e53362, 2013
- Sibanda N, McKenna A, Richmond A, Ricke SC, Callaway T, Stratakis ACh, Gundogdu O, Corcionivoschi N.** A Review of the Effect of Management Practices on *Campylobacter* Prevalence in Poultry Farms. *Front Microbiol.* 9, 2018
- Stefani S, Chung DR, Lindsay JA, Friedrich AW, Kearns AM, Westh H, MacKenzie FM.** Meticillin-resistant *Staphylococcus aureus* (MRSA): global epidemiology and harmonisation of typing methods. *Int. J. Antimicrob. Agents* 39, 273–282, 2012
- Toombs-Ruane LJ, Benschop J, Burgess S, Priest P, Murdoch DR, French NP.** Multidrug resistant *Enterobacteriaceae* in New Zealand: a current perspective. *N. Z. Vet. J.* 65, 62–70, 2017
- Türkyılmaz S, Tekbıyık S, Oryasin E, Bozdoğan B.** Molecular Epidemiology and Antimicrobial Resistance Mechanisms of Methicillin-Resistant *Staphylococcus aureus* Isolated from Bovine Milk. *Zoonoses and Public Health* 57, 197–203, 2010
- van Wagenberg CPA, van Horne PLM, Sommer HM, Nauta MJ.** Cost-effectiveness of *Campylobacter* interventions on broiler farms in six European countries. *Microb. Risk Anal.* 2–3, 53–62, 2016
- Welch D, Bansal S, Hunter DR.** Statistical inference to advance network models in epidemiology. *Epidemics* 3, 38–45, 2011

Williamson DA, Roberts SA, Ritchie SR, Coombs GW, Fraser JD, Heffernan H. Clinical and Molecular Epidemiology of Methicillin-Resistant *Staphylococcus aureus* in New Zealand: Rapid Emergence of Sequence Type 5 (ST5)-SCCmec-IV as the Dominant Community-Associated MRSA Clone. *PLoS One* 8, e62020, 2013

Supplementary Material
Poultry Industry Survey Questionnaire 2016

CONTACT STRUCTURES IN THE NEW ZEALAND COMMERCIAL POULTRY INDUSTRY



Dear Producer,

We are conducting a survey to describe the extent and nature of contacts between commercial farms in the New Zealand poultry industry. Specifically, we are interested in contacts made through the sale of feed, transportation of live poultry and poultry products, regular visits made by industry personnel, and the movement of waste products such as manure, litter and dead birds.

Why are we conducting this survey?

Contacts are an important means by which diseases can spread from farm to farm. There have been many changes in the poultry industry since the last contact survey was conducted in 2006. We want to know how these changes have impacted your risk of getting disease from other farms and how we can better help to protect your livelihoods.

How will we use the survey information?

We will use this data to reconstruct the industry contact network and develop simulation models to predict how diseases like Avian Influenza and Campylobacter may spread.

Confidentiality

All information will be treated as strictly confidential and no information will be used in any way that could reasonably be expected to identify individual persons, practices or organisations.

The study results will only be useful if **all** producers take part. We greatly appreciate your valuable time in completing the survey questions.

Sincerely,

Dr. Carolyn Gates
Massey University

Thank You!



Massey University



Ministry for Primary Industries
Manatū Ahu Matua



This research study is being conducted by the EpiCentre at Massey University and is fully endorsed by the Poultry Industry Association of New Zealand (PIANZ) and the Egg Producers Federation of New Zealand (EPF) with support from the Ministry for Primary Industries.

INSTRUCTIONS

Please read carefully before beginning the survey

In this questionnaire, we would like you to think about individuals and companies that have visited your farm (e.g. to supply or remove birds, goods or products) throughout the past year. We use the term **farm** to refer to facilities where birds or eggs are kept under a common system of management. In most situations, a farm will be a single physical location. Sometimes a farm may be made up of several distinct physical locations in close proximity, which we call **sites** (see Section 1.3).

For each **farm**, you will be asked to fill out information on:

Farm Demographics

1. Enterprise and farm details
2. Farm capacity and management

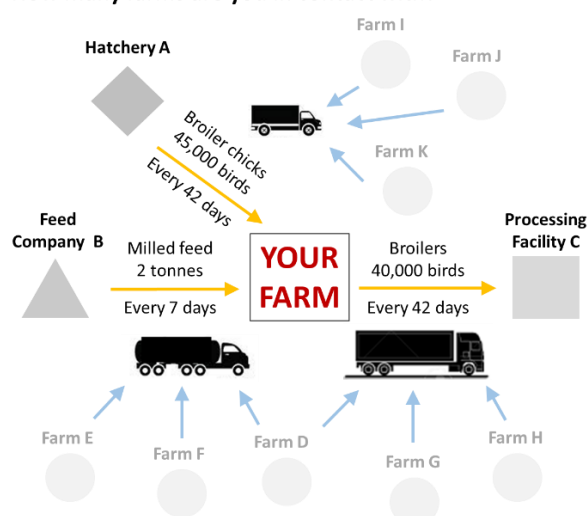
Routine Contacts

3. Transport companies
4. Feed companies
5. Live birds and hatching eggs
6. Table eggs and poultry products
7. Regular movements of personnel
8. Manure, litter, and dead birds

Other Information

9. Biosecurity risks
10. Additional feedback

How many farms are you in contact with?



We will need to link data from all survey respondents to create the final industry contact network so it is important that you provide us with as much accurate detail as possible about the identity of individuals and companies that have visited your farm. Once the study has been completed, you will be provided with a risk map similar to the figure above showing the number of direct (yellow) and indirect (blue) connections to your farm. The identity of the indirect connections will be hidden in the figure to preserve confidentiality.

Returning the questionnaire

Please return the questionnaire using the prepaid addressed envelope provided.

Questions

If you have any questions or concerns about completing the questionnaire or participating in the study, please contact:

Carolyn Gates

Massey University
 Institute of Veterinary, Animal and Biomedical Science (IVABS)
 Private Bag 11-222 Palmerston North 4442
 Tel: (06) 951 8140
 Email: c.gates@massey.ac.nz

1. ENTERPRISE AND FARM DETAILS

1.1. Details of the person filling in this questionnaire

Name	
Position	

1.2. Farm contact details

Name	
Street or PO Box number	
Town	
Phone (business)	
Phone (home)	
Phone (mobile)	
Facsimile	
Email	
Parent company (if applicable) ^a	

^a A parent company refers to the organisation that owns birds that are raised on one or more farms. Tegel Foods Limited, for example, would be referred to as a parent company.

1.3. Farm location details	Address (town or city only)	Agribase ^b (if present)	Farms Online ID ^c (if present)
Main site (1)			
Other site (2)			
Other site (3)			
Other site (4)			
Other site (5)			

^{b,c} We will only use this information to determine the spatial location of your farm so that we can report how many other commercial farms are in close proximity.

1.4. What type(s) of commercial poultry operation do you run on the farm?

If more than one type of commercial poultry operation is present, please indicate approximately what % each contributes to your total farm production.

	Present? (check all that apply)	% of total farm production
Commercial - Breeder	<input type="checkbox"/>	
Commercial - Layer hens	<input type="checkbox"/>	
Commercial - Pullets	<input type="checkbox"/>	
Commercial - Broilers	<input type="checkbox"/>	
Commercial - Turkeys	<input type="checkbox"/>	
Commercial - Hatchery	<input type="checkbox"/>	
Commercial - Ducks	<input type="checkbox"/>	
	Total	100%
Other (e.g. emu, ostrich, aviary birds)– please specify types and numbers of such birds.		

2. FARM CAPACITY AND MANAGEMENT

2.1. Over the past year, on average, how many live birds (or fertile eggs for hatcheries) were typically housed on the farm at any given time?

Please also indicate (a) the total number of sheds in which these birds were housed, (b) the average number of birds per shed, and (c) the total combined housing capacity of all sheds present on the farm.

Species	Approximate number of live birds (or fertile eggs)	(a) Total number of sheds	(b) Average shed capacity	(c) Total farm capacity
Layers				
Pullets				
Hatchery only (give fertile egg numbers)				
Broilers (when birds present)				
Turkeys				
Ducks				

2.2. How would you describe the typical flow of live birds (or fertile eggs for hatcheries) on and off farm?

Please indicate (a) if the management for the poultry species is all-in-all-out OR if multiple age classes are present at the same time, (b) the average length of the production cycle (#) specifying the **time units** in days, weeks, or months (i.e. 42 days), and (c) the average time that sheds are left empty before new birds (or fertile eggs) are introduced (#) specifying the **time units** in days, weeks, or months (i.e. 1 week).

Species	(a) Flow of birds (choose one)		(b) Length of production cycle		(c) Downtime between cycles	
	All-in-all-out	Multiple age classes	#	Time units	#	Time units
Layers	<input type="checkbox"/>	<input type="checkbox"/>				
Pullets	<input type="checkbox"/>	<input type="checkbox"/>				
Hatchery only (give fertile egg numbers)	<input type="checkbox"/>	<input type="checkbox"/>				
Broilers	<input type="checkbox"/>	<input type="checkbox"/>				
Turkeys	<input type="checkbox"/>	<input type="checkbox"/>				
Ducks	<input type="checkbox"/>	<input type="checkbox"/>				

2.3. Which of the following best describes your (a) current housing type and (b) intended housing type to comply with regulatory changes by 2022.

	(a) Current	(b) Future
Free range housing	<input type="checkbox"/>	<input type="checkbox"/>
Colony housing	<input type="checkbox"/>	<input type="checkbox"/>
Barn housing	<input type="checkbox"/>	<input type="checkbox"/>
I do not intend to produce after 2022		<input type="checkbox"/>

3. TRANSPORT COMPANIES

In this question, we are interested in knowing which transport companies and/or individuals have conveyed poultry or poultry related products (including eggs, feed, manure and dead birds) onto or off your farm during the past year.

For movements ONTO and OFF farm, please indicate the town or city of origin for the transporter (i.e. if Company A transported poultry from your farm to a processing plant and the company's depot was in Wanganui, list Wanganui as the address of origin of the transporter).

3.1. Please list the name and location details of individuals or companies that transported poultry or poultry-related product ONTO or OFF your farm over the last 12 months (in order of importance based on frequency of contacts). For part (d), please indicate the category number from the list below the table. For part (e), please indicate the average time interval between two consecutive contacts. For part (f), please indicate how much product is transferred with an average contact.

Example: Transport company X from Wellington transports 3000 adult birds onto your farm once every 40 days.

[illegible]

**** Categories include those who transport:**

- (1) Feed.
- (2) Hatching eggs.
- (3) Day old chicks, pullets, and adult birds.
- (4) Table eggs.
- (5) Poultry product.
- (6) Manure and litter.
- (7) Dead birds.
- (8) Broilers for slaughter

5. LIVE BIRDS AND HATCHING EGGS

In this question we are interested in the origin of live birds (including day-old chicks and point-of-lay pullets) and/or hatching eggs that came onto your farm and the destination of live birds (including broilers for processing, end-of-lay birds) and/or hatching eggs were sent from your farm to another location.

5.1. Please list the name and location details of individuals or companies that moved live birds and/or hatching eggs ONTO your farm or received live birds and/or hatching eggs moved OFF your farm over the last 12 months (in order of importance based on frequency of contacts). For part (d), please indicate the category number from the list below the table. For part (e), please indicate the average time interval between two consecutive contacts. For part (f), please indicate how much product is transferred with an average contact.
Example: 2500 adult broilers birds are moved to Processor Z in Hamilton from your farm every 42 days.

[illegible]

**** Categories include:**

- (1) Hatching eggs.
- (2) Day old chicks.
- (3) Pullets.
- (4) Adult birds.

7. REGULAR MOVEMENTS OF PERSONNEL

In this question, we are interested in the movement of people associated with the poultry industry onto your farm and the destination of permanent members of your farm staff to other locations where poultry are present. We are not interested in details of every single movement, only those that occur on a regular basis (e.g. routine visits by advisors, contractors, and personnel from other premises where poultry are kept).

7.1. Please list the name and location details of those permanent staff members who made an ON FARM or OFF FARM visit to other locations where poultry were present over the last 12 months (in order of importance). For part (a), please indicate the category number from the list below the table. For part (c), please indicate the average time interval between two consecutive contacts.

Example: A veterinarian based in Tauranga visits your farm for regular health inspections once every month.

[illegible]

**** Categories include:**

- (1) Veterinarians, advisors, industry representatives.
- (2) Contractors (those having direct contact with poultry sheds or equipment e.g. cleaning and maintenance crews).
- (3) Permanent employees on your farm
- (4) Individuals from premises where commercial poultry are kept.
- (5) Individuals from premises where non-commercial poultry are kept.

9. BIOSECURITY RISKS

9.1. Please answer the following questions about the workers on your farm.

How many full-time workers are present?

How many part-time workers are present?

Workers are assigned to (select one):

☐ Whole farm

☐ Specific sheds

9.2. Which of the following biosecurity measures are implemented on your property?

Please indicate how frequently they are implemented.

	Never	Rarely	Sometimes	Often	Always
Vehicle disinfection before entering farm	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dedicated clean coveralls for each shed	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dedicated boots or boot covers worn for each shed	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Footbaths at shed entrances	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Monitored rat and mouse bait stations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Bird-proofed housing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Bird-proofed feed storage	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

9.3. Which of the following equipment is shared with other farms?

Choose all that apply. For part (b), please indicate whether the shared equipment is cleaned and disinfected between farms.

	(a) Equipment	(b) Is the equipment cleaned and disinfected between farms?
Vehicles and trailers	<input type="checkbox"/>	<input type="checkbox"/> No <input type="checkbox"/> Yes
Gates and panels	<input type="checkbox"/>	<input type="checkbox"/> No <input type="checkbox"/> Yes
Lawn mowers	<input type="checkbox"/>	<input type="checkbox"/> No <input type="checkbox"/> Yes
Pressure sprayers / washers	<input type="checkbox"/>	<input type="checkbox"/> No <input type="checkbox"/> Yes
Skid steer loaders (tractors with buckets for cleaning out sheds)	<input type="checkbox"/>	<input type="checkbox"/> No <input type="checkbox"/> Yes
Other equipment	<input type="checkbox"/>	<input type="checkbox"/> No <input type="checkbox"/> Yes

9.4. Please answer the following questions about the water supply on your farm.

Select ONE option for each question

What is the main source of drinking water for poultry on your farm?	<input type="checkbox"/> Town supply <input type="checkbox"/> River <input type="checkbox"/> Bore <input type="checkbox"/> Other, please specify:
How is the drinking water treated?	<input type="checkbox"/> No treatment <input type="checkbox"/> Chlorine dioxide <input type="checkbox"/> Ultraviolet <input type="checkbox"/> Other, please specify:

9.5. Please answer the following questions about litter management on your farm.

What type of litter do you use?		
Is a litter shed present?	<input type="checkbox"/> No	<input type="checkbox"/> Yes
How often are the sheds cleaned out?	<input type="checkbox"/> Partially after each run	<input type="checkbox"/> Fully after each run
What disinfectant do you use?		
How moist is the litter?	<input type="checkbox"/> Too dry <input type="checkbox"/> About right <input type="checkbox"/> Too wet	
Do you use litter treatment?	<input type="checkbox"/> No <input type="checkbox"/> Yes, please specify:	

9.6. Do you see wild birds or waterfowl (a) in the same area as poultry (e.g. within sheds or within ranges on free-range operations) or (b) on any ponds or waterways present on the farm? If you do have ponds or waterways on the farm, (c) what is the approximate distance of the ponds or waterways from the poultry sheds?

	(a) In the same area as poultry?	(b) On ponds or waterways on the farm?	(c) Approximate distance of pond or waterway from sheds (metres)
Wild birds	<input type="checkbox"/> No <input type="checkbox"/> Yes	<input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> NA*	
Waterfowl	<input type="checkbox"/> No <input type="checkbox"/> Yes	<input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> NA*	

* I do not have any ponds or waterways present on the farm.

If you answered yes to question (a) and/or (b), please indicate which of the following wild bird or waterfowl species are observed:

	(a)	(b)
Sparrow	<input type="checkbox"/>	<input type="checkbox"/>
Starling	<input type="checkbox"/>	<input type="checkbox"/>
Sea gull	<input type="checkbox"/>	<input type="checkbox"/>
Duck	<input type="checkbox"/>	<input type="checkbox"/>
Others (please specify):	<input type="checkbox"/>	<input type="checkbox"/>

9.7. Please answer the following questions about contact with backyard (non-commercial) poultry operations.

	No	Yes	NA
Do point-of-lay birds get sold directly to backyard flocks from your farm?	<input type="checkbox"/>	<input type="checkbox"/> *	<input type="checkbox"/>
Do end-of-lay birds get sold directly to backyard flocks from your farm?	<input type="checkbox"/>	<input type="checkbox"/> *	<input type="checkbox"/>
How many independently-owned backyard flocks are adjacent to your farm (i.e share a farm boundary)?			

*** If you answered yes to either of these questions, would you mind telling us how these birds are advertised for sale?**

10. ADDITIONAL FEEDBACK**10.1. How concerned are you about the following poultry health issues on your farm?**

	Not at all concerned	Slightly concerned	Somewhat concerned	Moderately concerned	Extremely concerned
Campylobacter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Salmonella	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Avian influenza	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Coccidiosis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Antibiotic resistance	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Welfare	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Comments:

10.2. What would you describe as the biggest concern facing your farm?

10.3. What areas of future research would have the most benefit to your farm?

10.4. Would you be willing to participate in follow-up surveys?

Yes	<input type="checkbox"/>
No	<input type="checkbox"/>

Thank You!

We greatly appreciate your time in filling out this survey.

**Massey University**Ministry for Primary Industries
Manatū Ahu Matua

Supplementary Material
Chapter 3

B.1. Network reconstruction

In order to calculate a degree centrality score for each of the study farms, several network graphs were constructed from the reported on- and off-farm movement of goods and services. To begin with, bimodal networks were built with nodes representing either a study farm, a transporting company or a group of personnel. Transporting companies were identified from the survey responses with companies responsible for the on- and off-farm movement of either (i) feed, (ii) waste and litter, (iii) live birds and hatching eggs or (iv) any other poultry product. Before constructing the network graphs, the company names were cross-checked for variation and spelling errors to ensure that the same company was not listed multiple times. Where company descriptions were unclear, clarification was sought from PIANZ and/or MPI staff. If single companies had enterprises in multiple locations, each separate location was assigned a unique identification number by name and address so it would appear as a unique node in the network analysis. For the reported movement of personnel, neither individuals, or the companies they work for, could be identified from the survey responses, however, personnel had been categorised within one of five groups; (i) an employee, (ii) a contractor, (iii) an individual in contact with commercial poultry, (iv) an individual in contact with non-commercial poultry or (v) a veterinarian, advisor or industry representative. These categorises were used to create additional network nodes by subdividing personnel within each category by the district they originate from, as reported in the survey. For example, all veterinarians, advisors or industry representatives from the Manawatu region would form one group which could then be used as a node in the network in addition to transporting companies and study farms.

Network graphs could then be constructed by forming an undirected edge between each study farm and every transporting company or personnel group that they had reported in the survey. Each edge had information attached regarding the movement frequency; that is the number of days in between two consecutive movements, and the quantity of product moved. All the numeric variables describing the frequency and quantity of items transferred between the study farms were checked for conflicts in the unit of measure and standardised as needed. All together six bimodal networks were constructed such that each network graph showed just the movements relating to either (i) feed, (ii) waste and

litter, (iii) live birds and hatching eggs, (iv) all other poultry products, (v) personnel or (vi) all the reported movements combined.

Using each of the bimodal networks, six additional unimodal network graphs were also constructed with nodes representing only the study farms. Unimodal network graphs were built by forming an undirected edge between each of the study farms in the network that shared a link to a common transporting company or personnel group (Figure B3). Both the bimodal and unimodal network graphs were plotted using a force-based algorithm proposed by Fruchterman and Reingold (1991) to help visualise the network structure (Figure B4). In each graph, any study farm that did not report a movement within one of the networks or reported internal movements (*e.g.*, spreading litter on-site) can still be seen as isolated nodes without edges. Basic network statistics were calculated to describe each network graph in terms of their overall size, the frequency of movements and the quantity of products moved. In addition, the degree centrality and betweenness were also calculated for each of the study farms using the unimodal network graphs to identify individuals with the greatest number of on- and off-farm movements (degree) and individuals most frequently found on the shortest path between two other farms in the network (betweenness) (Table B9). The calculated degree measure from the combined unimodal network graph, showing all on- and off-farm movements, was used as the risk criterion in the main study analysis (Table B10). Lastly, degree distributions were plotted to distinguish any major network structures in comparison to other real-world networks (Figure B1).

Table B1. Frequency of implementing seven common biosecurity measures amongst 120 producers in the New Zealand commercial poultry industry including 33 layer enterprises (including mixed pullet and layer operations), 57 broiler enterprises, 24 breeder enterprises and 6 other poultry enterprises (including duck, turkey and pullet operations).

		Number of farms (%)			
		Layers (n = 33)	Broilers (n = 57)	Breeders (n = 24)	Other poultry (n = 6)
Vehicle disinfection	<i>Never</i>	30 (90.9)	20 (35.1)	4 (16.7)	2 (33.3)
	<i>Rarely</i>	1 (3.0)	5 (8.8)	8 (33.3)	1 (16.7)
	<i>Sometimes</i>	2 (6.1)	10 (17.5)	1 (4.2)	0 (0.0)
	<i>Often</i>	0 (0.0)	8 (14.0)	0 (0.0)	1 (16.7)
	<i>Always</i>	0 (0.0)	14 (24.6)	11 (45.8)	2 (33.3)
Clean coveralls	<i>Never</i>	20 (60.6)	18 (31.6)	1 (4.2)	6 (100)
	<i>Rarely</i>	2 (6.1)	3 (5.3)	0 (0.0)	0 (0.0)
	<i>Sometimes</i>	4 (12.1)	15 (26.3)	1 (4.2)	0 (0.0)
	<i>Often</i>	3 (9.1)	4 (7.0)	0 (0.0)	0 (0.0)
	<i>Always</i>	4 (12.1)	17 (29.8)	22 (91.7)	0 (0.0)
Boot covers	<i>Never</i>	18 (54.5)	1 (1.8)	1 (4.2)	1 (16.7)
	<i>Rarely</i>	1 (3.0)	0 (0.0)	0 (0.0)	0 (0.0)
	<i>Sometimes</i>	3 (9.1)	2 (3.5)	0 (0.0)	0 (0.0)
	<i>Often</i>	2 (6.1)	0 (0.0)	0 (0.0)	3 (50.0)
	<i>Always</i>	9 (27.3)	54 (94.7)	23 (95.8)	2 (33.3)
Footbaths	<i>Never</i>	21 (63.6)	10 (17.5)	0 (0.0)	0 (0.0)
	<i>Rarely</i>	0 (0.0)	2 (3.5)	11 (45.8)	0 (0.0)
	<i>Sometimes</i>	2 (6.1)	4 (7.0)	0 (0.0)	0 (0.0)
	<i>Often</i>	2 (6.1)	2 (3.5)	0 (0.0)	0 (0.0)
	<i>Always</i>	8 (24.2)	39 (68.4)	13 (54.2)	6 (100)
Rodent bait stations	<i>Never</i>	1 (3.0)	0 (0.0)	0 (0.0)	0 (0.0)
	<i>Rarely</i>	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	<i>Sometimes</i>	1 (3.0)	0 (0.0)	0 (0.0)	0 (0.0)
	<i>Often</i>	1 (3.0)	2 (3.5)	0 (0.0)	0 (0.0)
	<i>Always</i>	30 (90.9)	55 (96.5)	24 (100)	6 (100)

Table B1 continues next page

Table B1 continued

Bird-proofed housing	<i>Never</i>	7 (21.2)	1 (1.8)	0 (0.0)	0.(0.0)
	<i>Rarely</i>	3 (9.1)	0 (0.0)	0 (0.0)	0 (0.0)
	<i>Sometimes</i>	1 (3.0)	4 (7.0)	0 (0.0)	0 (0.0)
	<i>Often</i>	2 (6.1)	0 (0.0)	0 (0.0)	0 (0.0)
	<i>Always</i>	20 (60.6)	52 (91.2)	24 (100)	6 (100)
Bird-proofed feed store	<i>Never</i>	3 (9.1)	0 (0.0)	1 (4.2)	0.(0.0)
	<i>Rarely</i>	0 (0.0)	0.(0.0)	0.(0.0)	0 (0.0)
	<i>Sometimes</i>	2 (6.1)	0.(0.0)	0.(0.0)	0 (0.0)
	<i>Often</i>	1 (3.0)	1 (1.8)	0.(0.0)	0 (0.0)
	<i>Always</i>	27 (81.8)	56 (98.2)	23 (95.8)	6 (100)

Table B2. Results from multiple Dunn's tests comparing the estimated biosecurity score between poultry production types following a significant Kruskal-Wallis test ($\chi^2 = 43.99$, $df = 3$, $p < 0.0001$). Significance values are based on the adjusted p -value, adjusted for multiple testing using the Bonferroni correction.

	<i>z</i> -score	Adjusted <i>p</i> -value
Breeder – Broiler	1.43	0.4572
Breeder – Layer ***	5.78	<0.0001
Broiler – Layer ***	5.49	<0.0001
Breeder – Other *	2.89	0.0114
Broiler – Other	2.21	0.0821
Layer – Other	-0.99	0.9700
Level of significance *** $p < 0.001$, ** $p < 0.001$, * $p < 0.1$		

Table B3. Results from multiple Dunn's tests comparing the estimated disease contact risk score between poultry production types following a significant Kruskal-Wallis test ($\chi^2 = 25.48$, $df = 3$, $p < 0.0001$). Significance values based on the adjusted p-value, adjusted for multiple testing using the Bonferroni correction.

	<i>z</i>-score	Adjusted <i>p</i>-value
Breeder - Broiler *	-3.33	0.0026
Breeder - Layer ***	-4.72	<0.0001
Broiler - Layer	-2.12	0.1008
Breeder - Other *	-3.46	0.0018
Broiler - Other	-1.56	0.3572
Layer - Other	-0.29	1.00
Level of significance *** $p < 0.001$, ** $p < 0.01$, * $p < 0.1$		

Table B4. The breakdown of 120 producers in the New Zealand commercial poultry industry, including 33 layer enterprises (including mixed pullet and layer operations), 57 broiler enterprises, 24 breeder enterprises and 6 other poultry enterprises (including duck, turkey and pullet operations), and the number of neighbouring commercial poultry farms within a 5km radius of their production premises.

Number of neighbouring farms	Number of farms (% within the sector)			
	Layers (n = 33)	Broilers (n = 57)	Breeders (n = 24)	Other poultry (n = 6)
1	18 (54.5)	6 (10.5)	2 (8.3)	3 (50.0)
2	6 (18.2)	6 (10.5)	8 (33.3)	1 (16.7)
3	4 (12.1)	6 (10.5)	0 (0.0)	0 (0.0)
4	1 (3.3)	7 (12.3)	3 (12.5)	0 (0.0)
5	0 (0.0)	5 (8.8)	2 (8.3)	0 (0.0)
6	1 (3.3)	3 (5.3)	3 (12.5)	0 (0.0)
7	0 (0.0)	2 (3.5)	5 (20.8)	2 (33.3)
≥ 8	3 (9.9)	22 (38.6)	1 (4.2)	0 (0.0)

Table B5. Water source and treatment on 120 producers in the New Zealand commercial poultry industry, including 33 layer operations (including mixed pullet and layer enterprises), 57 broiler enterprises, 24 breeder enterprises and 6 other poultry enterprises (including duck, turkey and pullet operations).

		Number of farms (% within the sector)			
		Layers (n = 33)	Broilers (n = 57)	Breeders (n = 24)	Other poultry (n = 6)
Water source	<i>Town supply</i>	6 (18.2)	10 (17.5)	18 (75.0)	3 (50.0)
	<i>River/ stream/ spring</i>	3 (9.1)	4 (7.0)	0 (0.0)	0 (0.0)
	<i>Bore</i>	16 (48.5)	43 (75.4)	6 (25.0)	3 (50.0)
	<i>Rain/ roof water</i>	8 (24.2)	0 (0.0)	0 (0.0)	0 (0.0)
Treatment	<i>No treatment</i>	16 (48.5)	3 (5.3)	4 (16.7)	5 (83.3)
	<i>Chlorine Dioxide</i>	7 (21.2)	52 (91.2)	9 (37.5)	0 (0.0)
	<i>Ultraviolet</i>	8 (24.2)	0 (0.0)	0 (0.0)	0 (0.0)
	<i>Sodium Hypochlorite</i>	0 (0.0)	2 (3.5)	11 (45.8)	0 (0.0)
	<i>Filtration</i>	2 (6.1)	0 (0.0)	0 (0.0)	0 (0.0)
	<i>Citric acid mix</i>	0 (0.0)	0 (0.0)	0 (0.0)	1 (16.7)

Table B6. Litter management on 120 producers in the New Zealand commercial poultry industry, including 33 layer operations (including mixed pullet and layer enterprises), 57 broiler enterprises, 24 breeder enterprises and 6 other poultry enterprises (including duck, turkey and pullet operations).

		Number of farms (% within the sector)			
		Layers (n = 33)	Broilers (n = 57)	Breeders (n = 24)	Other poultry (n = 6)
Litter type	<i>No litter</i>	15 (45.5)	6 (10.5)	2 (8.3)	2 (33.3)
	<i>Sawdust</i>	5 (15.2)	0 (0.0)	0 (0.0)	0 (0.0)
	<i>Wood shavings</i>	10 (30.3)	51 (89.5)	0 (0.0)	0 (0.0)
	<i>Other</i>	3 (9.09)	0 (0.0)	22 (91.7)	4 (66.7)
Litter shed	<i>Present</i>	10 (30.3)	4 (7.0)	2 (8.3)	4 (66.7)
	<i>Absent</i>	23 (69.7)	53 (93.0)	22 (91.7)	2 (33.3)
Regularity of cleaning sheds	<i>Partially</i>	3 (9.09)	0 (0.0)	0 (0.0)	2 (33.3)
	<i>Fully</i>	23 (69.7)	57 (100)	23 (95.8)	4 (66.7)
	<i>N/A</i>	7 (21.2)	0 (0.0)	1 (4.2)	0 (0.0)
Disinfectant used	<i>None</i>	14 (42.4)	19 (33.3)	3 (12.5)	3 (50.0)
	<i>Techsan</i>	0 (0.0)	4 (7.0)	3 (12.5)	1 (16.7)
	<i>Glutasan</i>	0 (0.0)	4 (7.0)	11 (45.8)	0 (0.0)
	<i>Enviro-san</i>	2 (6.06)	5 (8.8)	0 (0.0)	0 (0.0)
	<i>Virkon</i>	9 (27.3)	8 (14.0)	0 (0.0)	1 (16.7)
	<i>Other</i>	8 (24.2)	17 (29.8)	7 (29.2)	1 (16.7)
Litter moisture	<i>Too dry</i>	2 (6.06)	1 (1.8)	0 (0.0)	0 (0.0)
	<i>About right</i>	24 (72.7)	56 (98.2)	23 (95.8)	6 (100)
	<i>Too wet</i>	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	<i>N/A</i>	7 (21.2)	0 (0.0)	1 (4.2)	0 (0.0)
Litter treatment	<i>No</i>	24 (72.7)	57 (100)	23 (95.8)	6 (100)
	<i>Yes</i>	2 (6.06)	0 (0.0)	0 (0.0)	0 (0.0)
	<i>N/A</i>	7 (21.2)	0 (0.0)	1 (4.2)	0 (0.0)

Table B7. Reported sightings of wild birds, including six common species and waterfowl, on 120 producers in the New Zealand commercial poultry industry, including 33 layer enterprises (including mixed pullet and layer enterprises), 57 broiler enterprises, 24 breeder enterprises and 6 other poultry enterprises (including duck, turkey and pullet operations).

		Number of farms (% within the sector)			
		Layers (n = 33)	Broilers (n = 57)	Breeders (n = 24)	Other poultry (n = 6)
In the same area as poultry (%)	<i>Wild birds</i>	21 (63.6)	20 (35.1)	1 (4.2)	3 (50.0)
	<i>Waterfowl</i>	3 (9.1)	5 (8.8)	1 (4.2)	0 (0.0)
	<i>Sparrows</i>	21 (63.6)	24 (42.1)	0 (0.0)	3 (50.0)
	<i>Starlings</i>	15 (45.5)	14 (24.6)	0 (0.0)	3 (50.0)
	<i>Sea gulls</i>	4 (12.1)	5 (8.8)	0 (0.0)	0 (0.0)
	<i>Duck</i>	3 (9.1)	7 (12.3)	1 (4.2)	1 (16.7)
	<i>Other</i>	6 (18.2)	8 (14.0)	8 (33.3)	2 (33.3)
On ponds or waterways on the farm (%)	<i>Wild birds</i>	7 (21.2)	14 (24.6)	1 (4.2)	3 (50.0)
	<i>Waterfowl</i>	7 (21.2)	17 (29.8)	6 (25.0)	3 (50.0)
	<i>Sparrows</i>	2 (6.1)	10 (17.5)	0 (0.0)	1 (16.7)
	<i>Starlings</i>	3 (9.1)	7 (12.3)	0 (0.0)	1 (16.7)
	<i>Sea gulls</i>	3 (9.1)	1 (1.8)	0 (0.0)	0 (0.0)
	<i>Duck</i>	7 (21.2)	15 (26.3)	1 (4.2)	1 (16.7)
	<i>Other</i>	1 (3.0)	5 (8.8)	1 (4.2)	1 (16.7)

Table B8. Equipment shared by 120 producers in the New Zealand commercial poultry industry and the proportion cleaning them upon return.

	Number of farms sharing equipment (n = 120) (% cleaned)	% of farms that shared but did not clean equipment	Number of farms not sharing equipment (n = 120) (%)
Vehicles	39 (53.8, 21/39)	46.2 (18/39)	81 (67.5)
Gates	21 (33.3, 7/21)	66.7 (14/21)	99 (82.5)
Lawn equipment	33 (51.5, 17/33)	48.5 (16/33)	87 (72.5)
Sprayers	47 (63.8, 30/47)	36.2 (17/47)	73 (60.8)
Skid steer loaders	53 (67.9, 36/53)	32.1 (17/53)	67 (55.8)
Other equipment	43 (65.1, 28/43)	34.9 (15/43)	77 (64.1)

Table B9. Network statistics for the five contact networks constructed from the on- and off-farm movements of (i) feed, (ii) litter and waste, (iii) live birds and hatching eggs, (iv) personnel and, (v) all other poultry products across 120 producers in the New Zealand commercial poultry industry. Both the frequency (*i.e.*, the number of days between two consecutive contacts on the same farm) and quantity (*i.e.*, the amount of goods transferred in each movement with varying units across the contact networks) have been given in addition to the network degree (*i.e.*, the total number of movements onto and off a farm) and betweenness (*i.e.*, the frequency a farm is in the shortest path between two other farms in the network).

	Feed	Litter and waste	Live birds & hatching eggs	Personnel	Poultry products
Number of poultry premises ^a	115	112	117	92	45
Number of unique companies ^b	23	87	38	49	50
Total number of nodes	138	199	155	141	95
Movements onto-farm	5305	1246	3512	NA ^c	30
Movements off-farm	148	2337	3932	NA ^c	130
Total number of edges	5453	3583	7444	4462	160
Mean frequency (min-max)	13 (1-270)	68 (1-450)	263 (1-18250)	20533 (1-799350)	9 (1-100)
Mean quantity (min-max)	5101 tonnes (0.2-300000)	107 tonnes (0.01-4000)	54847 birds (210-413000)	1 person (NA)	1130 dozen (8-7000)
Mean degree ^c (min-max, median)	20.3 (0-42, 28)	11.7 (0-37,14)	20.6 (0-45, 26)	13.0 (0-43, 16)	1.1 (0-10, 0)
Mean betweenness ^d (min-max, median)	64.0 (0-2070, 0.0)	63.4 (0-1056, 1.3)	59 (0-551, 2.4)	46.6 (0-1186, 0.0)	0.1 (0-10, 0.0)

^a Poultry premises with degree>0

^b Companies are those providing goods and services to poultry premises in the network. If single companies had enterprises in multiple locations, each separate location was assigned a unique identification number by name and address so it would appear as a unique company in the network analysis

^c Movement of personnel considered undirected

Table B10. Basic network statistics for the contact network constructed from the on- and off-farm movements of all goods and services across 120 producers in the New Zealand commercial poultry industry including 33 layer enterprises (including mixed pullet and layer operations), 57 broiler enterprises, 24 breeder enterprises and 6 other poultry enterprises (including duck, turkey and pullet operations). Definitions for each measure areas follows; “*degree*” = the total number of movements onto and off a farm, “*betweenness*” = the frequency a farm is in the shortest path between two other farms in the network, “*network density*” = the proportion of all possible links between farms in the network that are actually present, “*average path length*” = the average shortest path between any pair of farms in the network averaged over all pairs of farms, “*clustering coefficient*” = for any farm in the network the clustering coefficient is the proportion of neighbouring farms in direct contact with the farms that are also connected to each other, “*network diameter*” = the longest path between any two pair of farms in the network and, “*fragmentation*” = the proportion of farm pairs for which a path does not exist between them.

Network statistic	Network nodes	
Mean degree (min-max, median)	<i>Layers (n = 33)</i>	13.0 (0-41, 9.0)
	<i>Broilers (n = 57)</i>	29.6 (8-57, 29.0)
	<i>Breeders (n =24)</i>	35.5 (22-50, 34.0)
	<i>Other poultry (n = 6)</i>	29.7 (4-37, 34.0)
	<i>All nodes (n = 120)</i>	26.2 (0-57, 29.0)
Mean betweenness (min-max, median)	<i>Layers (n = 33)</i>	92.0 (0-640, 33.8)
	<i>Broilers (n = 57)</i>	59.7 (0-650, 4.8)
	<i>Breeders (n =24)</i>	52.7 (0-223, 42.7)
	<i>Other poultry (n = 6)</i>	132.9 (0-545, 0.9)
	<i>All nodes (n = 120)</i>	70.9 (0-650, 20.6)
Network density	<i>All nodes (n = 120)</i>	0.220
Average path length	<i>All nodes (n = 120)</i>	2.211
Clustering coefficient	<i>All nodes (n = 120)</i>	0.777
Network diameter	<i>All nodes (n = 120)</i>	6
Fragmentation	<i>All nodes (n = 120)</i>	0.017

Figure B1. Degree distribution for the six contact networks constructed from the on- and off-farm movements of (a) all good and services, (b) feed, (c) waste and litter, (d) live birds and hatching eggs, (e) table eggs and poultry products and, (f) personnel report by 120 producers within the New Zealand commercial poultry industry. Graphs include the mean degree (μ) and median degree (\tilde{x}) for each network.

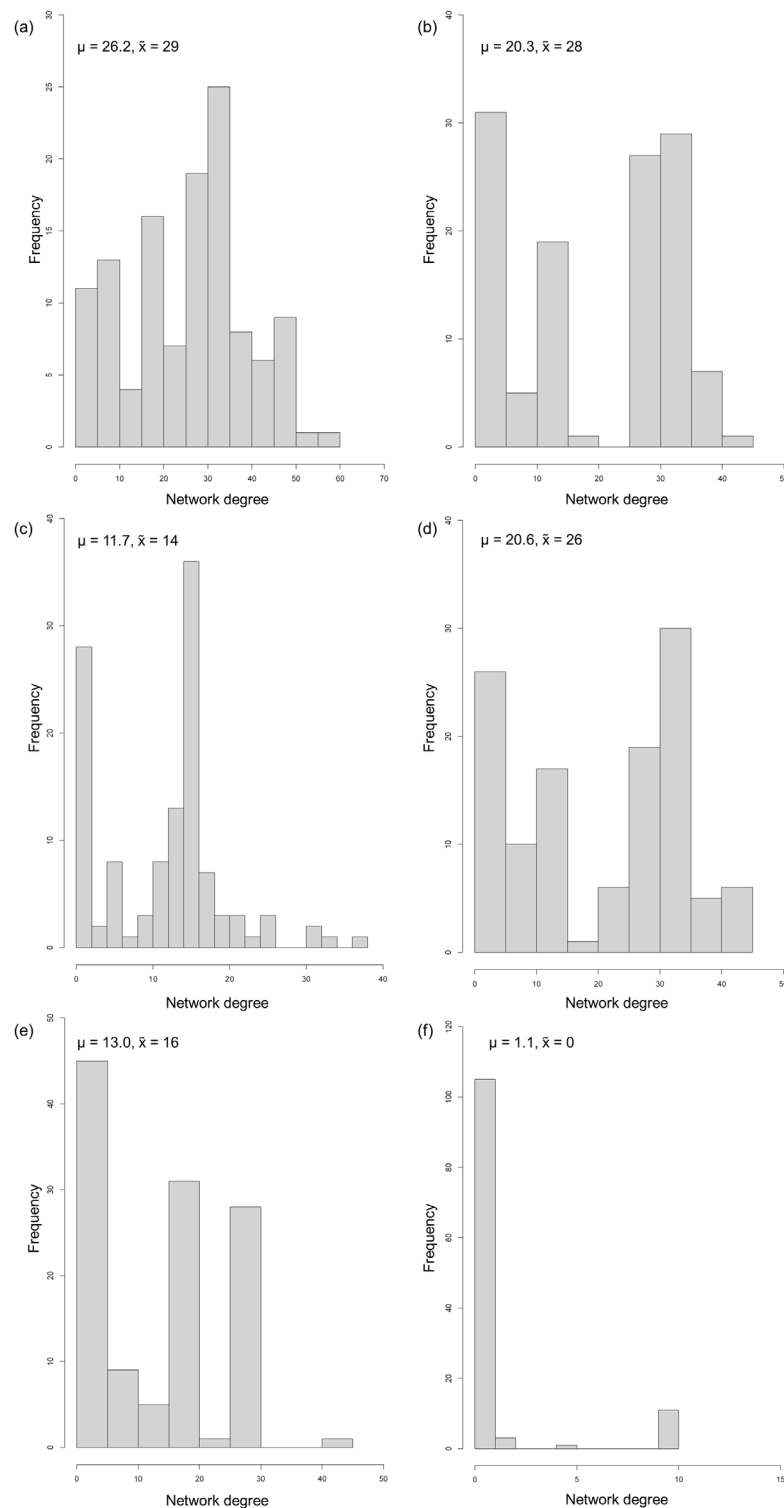


Figure B2. Scatter plots showing the relationship between the estimated biosecurity score, calculated from the reported frequency of implementing seven common biosecurity practices, and the estimated disease contact risk score across (a) 57 broiler enterprises, (b) 33 layer enterprises (including mixed pullet and layer operations), (c) 24 breeder enterprises and (d) 6 other poultry enterprises (including duck, turkey or pullet operations) in the New Zealand commercial poultry industry. The Pearson’s correlation coefficient (PCC) and 95% confidence intervals were -0.03 (-0.29, 0.23), -0.02 (-0.11, 0.33), 0.66 (0.34, 0.84), and -0.01 (-0.82, 0.81) for plots (a), (b), (c) and (d) respectively (p -value <0.01).

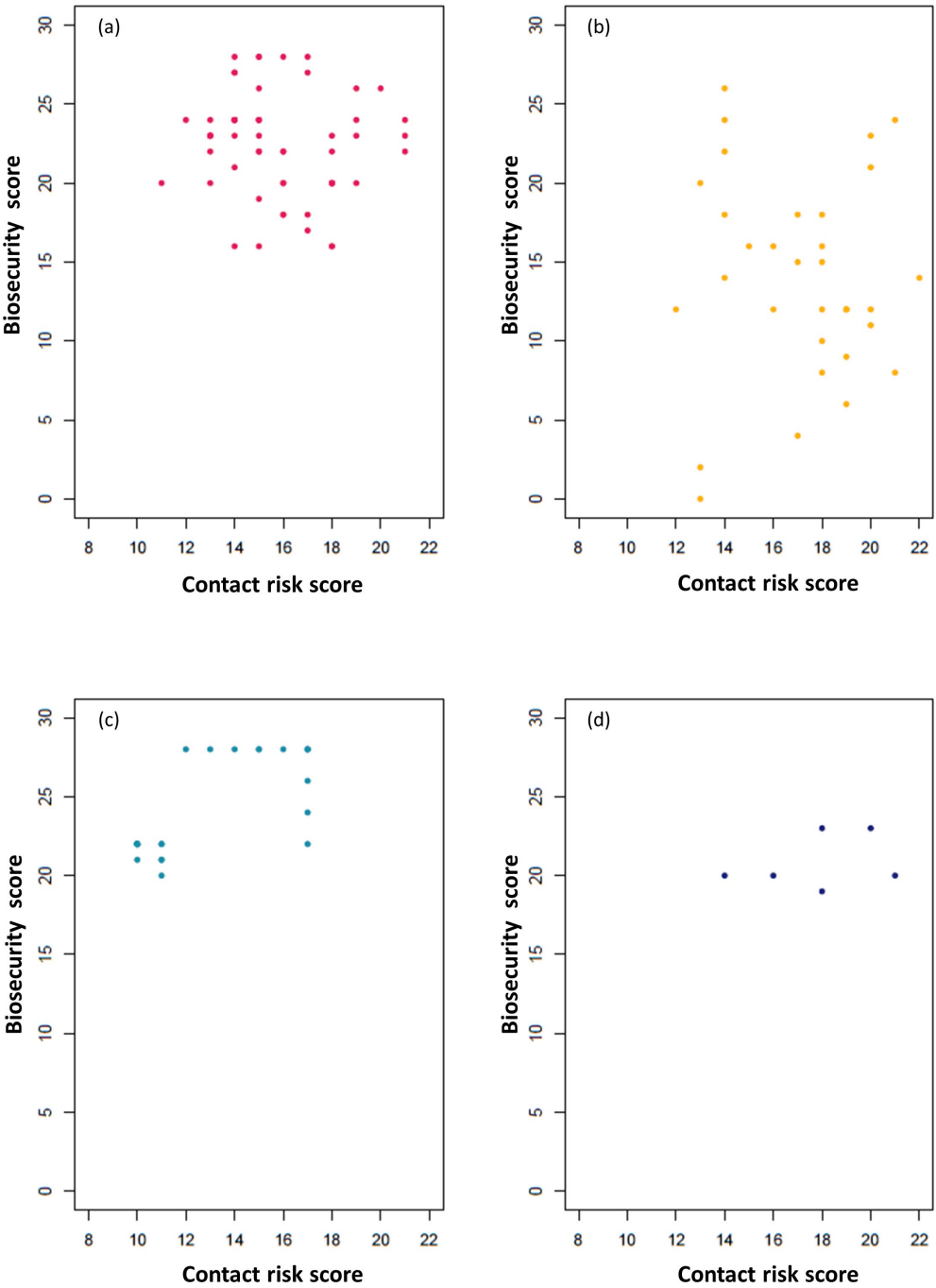
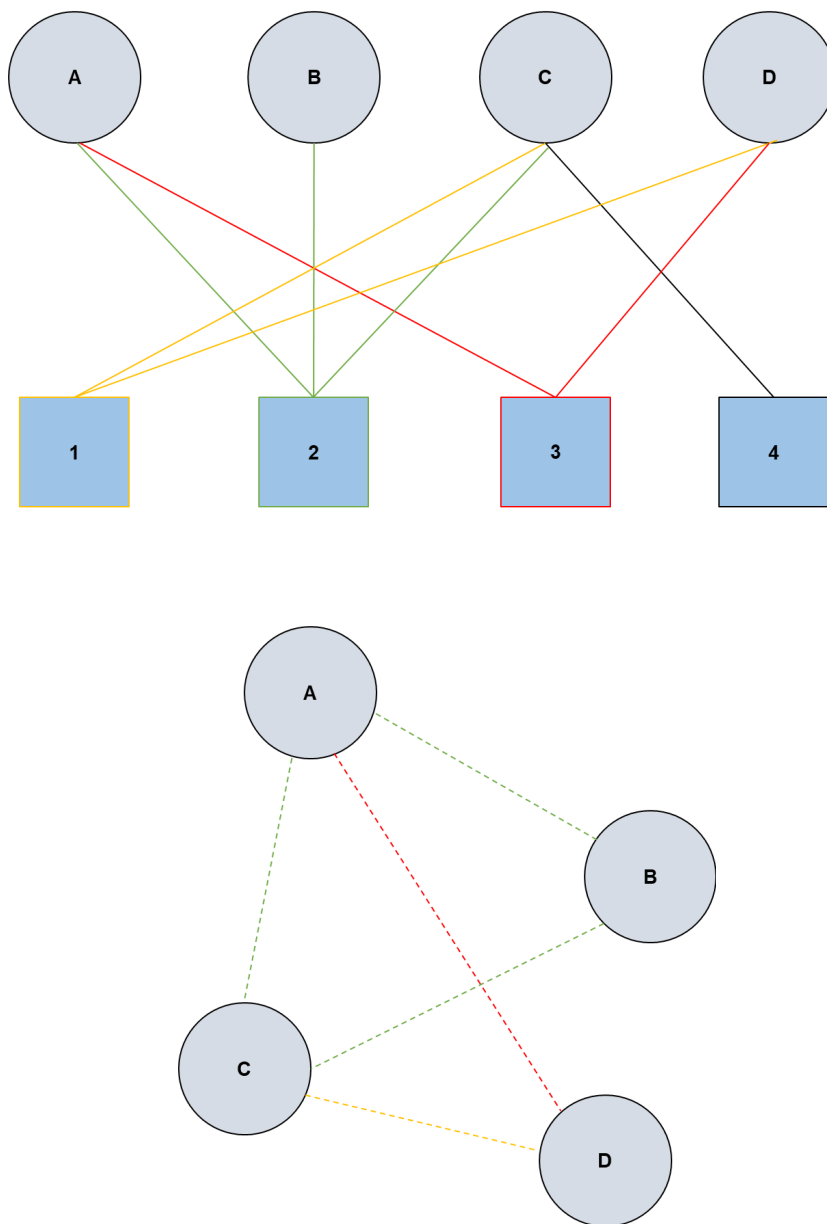


Figure B3. Bimodal networks (BMN) have vertices (V) belonging to different groups or modes. The top graph shows a schematic of the study BMN; V1 (A, B, C, and D) are poultry enterprises and V2 (1, 2, 3 and 4) are those companies providing goods and services to each operation. The bottom graph shows the unimodal network (UMN) constructed from BMN above by forming an edge between vertices belonging to V1 if they share common vertices belonging to V2 i.e. A is connected to D via their shared connection to 3. Edge colour in both graphs corresponds to the movement of different goods and services in the networks.



Supplementary Material
Chapter 4

Table C1. Summary network statistics for the trader networks showing poultry trades occurring through the online auction website TradeMe® in New Zealand from 01st January 2012 to 31st December 2017. For each year four trader networks were constructed showing the transaction of (i) chickens, (ii) bantams, (iii) ducks and, (iv) all poultry. Measures include the number of nodes, number of links (*i.e.*, edges), network density, diameter, average path length, clustering coefficient, number of giant strongly connected components (GSCC) and giant weakly connected components (GWCC), reciprocity and, fragmentation.

Trader network		2012	2013	2014	2015	2016	2017
Number of nodes	<i>Chickens</i>	11643	12405	11686	12083	11848	11807
	<i>Bantams</i>	1977	1874	1711	1537	1467	1391
	<i>Ducks</i>	1469	1622	1688	1757	1723	1697
	<i>All poultry</i>	13220	14049	13360	13737	13461	13422
Number of links	<i>Chickens</i>	15368	15787	14785	15172	14687	14313
	<i>Bantams</i>	2188	1940	1721	1512	1465	1347
	<i>Ducks</i>	1396	1508	1662	1675	1622	1540
	<i>All poultry</i>	18848	19134	18064	18260	17687	17109
Density	<i>Chickens</i>	1.13x10 ⁻⁴	1.03x10 ⁻⁴	1.08x10 ⁻⁴	1.04x10 ⁻⁴	1.05x10 ⁻⁴	1.03x10 ⁻⁴
	<i>Bantams</i>	5.60x10 ⁻⁴	5.53x10 ⁻⁴	5.88x10 ⁻⁴	6.40x10 ⁻⁴	6.81x10 ⁻⁴	6.97x10 ⁻⁴
	<i>Ducks</i>	6.47x10 ⁻⁴	5.74x10 ⁻⁴	5.84x10 ⁻⁴	5.43x10 ⁻⁴	5.47x10 ⁻⁴	5.35x10 ⁻⁴
	<i>All poultry</i>	1.08x10 ⁻⁴	9.69x10 ⁻⁵	1.01x10 ⁻⁴	9.68x10 ⁻⁵	9.76x10 ⁻⁵	9.50x10 ⁻⁵
Diameter	<i>Chickens</i>	24	19	19	19	19	22
	<i>Bantams</i>	16	13	8	9	10	7
	<i>Ducks</i>	6	5	6	5	7	5
	<i>All poultry</i>	19	19	19	19	17	18
Average path length	<i>Chickens</i>	8.85	7.12	6.69	6.63	7.33	6.96
	<i>Bantams</i>	4.84	4.61	2.87	2.51	2.56	2.27
	<i>Ducks</i>	1.67	1.39	1.61	1.74	1.52	1.42
	<i>All poultry</i>	7.10	6.91	6.70	6.49	7.03	6.66
Clustering coefficient	<i>Chickens</i>	2.71x10 ⁻³	2.94x10 ⁻³	3.35x10 ⁻³	3.78x10 ⁻³	2.28x10 ⁻³	1.08x10 ⁻³
	<i>Bantams</i>	1.37x10 ⁻²	9.57x10 ⁻³	7.03x10 ⁻³	4.29x10 ⁻³	4.35x10 ⁻³	5.63x10 ⁻³
	<i>Ducks</i>	5.96x10 ⁻³	3.41x10 ⁻³	3.68x10 ⁻³	1.56x10 ⁻³	2.68x10 ⁻³	1.60x10 ⁻³
	<i>All poultry</i>	4.15x10 ⁻³	3.79x10 ⁻³	4.04x10 ⁻³	4.21x10 ⁻³	3.12x10 ⁻³	1.39x10 ⁻³
Number nodes in GSCC	<i>Chickens</i>	218	242	215	213	141	79
	<i>Bantams</i>	24	9	2	3	3	5
	<i>Ducks</i>	1	1	1	2	1	1
	<i>All poultry</i>	393	351	321	299	226	94

Table C1 continues next page

Table C1 continued

Number nodes in GWCC	<i>Chickens</i>	10647	11114	10423	10767	10458	10442
	<i>Bantams</i>	1512	1385	1254	1044	1049	874
	<i>Ducks</i>	956	1009	1109	1165	1014	1005
	<i>All poultry</i>	12230	12801	12120	12405	12116	12031
Reciprocity	<i>Chickens</i>	2.86×10^{-3}	1.90×10^{-3}	2.71×10^{-3}	2.24×10^{-3}	8.17×10^{-4}	8.38×10^{-4}
	<i>Bantams</i>	1.83×10^{-3}	3.09×10^{-3}	1.16×10^{-3}	1.32×10^{-3}	1.37×10^{-3}	1.48×10^{-3}
	<i>Ducks</i>	0.0	0.0	0.0	1.19×10^{-3}	0.0	0.0
	<i>All poultry</i>	3.40×10^{-3}	2.72×10^{-3}	3.32×10^{-3}	2.63×10^{-3}	1.02×10^{-3}	1.05×10^{-3}
Fragmentation	<i>Chickens</i>	1.0	1.0	1.0	1.0	1.0	1.0
	<i>Bantams</i>	1.0	1.0	1.0	1.0	1.0	1.0
	<i>Ducks</i>	1.0	1.0	1.0	1.0	1.0	1.0
	<i>All poultry</i>	1.0	1.0	1.0	1.0	1.0	1.0

Table C2. Summary network statistics for the spatial networks showing poultry trades occurring through the online auction website TradeMe® in New Zealand from 01st January 2012 to 31st December 2017. For each year four trader networks were constructed showing the transaction of (i) chickens, (ii) bantams, (iii) ducks and, (iv) all poultry. Measures include the number of nodes, number of links (*i.e.*, edges), network density, diameter, average path length, clustering coefficient, number of giant strongly connected components (GSCC) and giant weakly connected components (GWCC), reciprocity and, fragmentation.

Spatial Network		2012	2013	2014	2015	2016	2017
Number of nodes	<i>Chickens</i>	130	131	129	130	133	134
	<i>Bantams</i>	122	119	122	115	115	117
	<i>Ducks</i>	119	119	115	116	121	112
	<i>All poultry</i>	130	131	130	130	133	134
Number of links	<i>Chickens</i>	2528	2456	2317	2486	2413	2401
	<i>Bantams</i>	853	798	713	651	655	594
	<i>Ducks</i>	642	642	647	612	682	643
	<i>All poultry</i>	2992	2874	2697	2824	2786	2783
Density	<i>Chickens</i>	0.15	0.14	0.14	0.15	0.14	0.14
	<i>Bantams</i>	0.06	0.06	0.05	0.05	0.05	0.04
	<i>Ducks</i>	0.05	0.05	0.05	0.05	0.05	0.05
	<i>All poultry</i>	0.18	0.17	0.16	0.17	0.16	0.16
Diameter	<i>Chickens</i>	5	5	5	4	5	5
	<i>Bantams</i>	7	7	8	6	7	7
	<i>Ducks</i>	6	6	7	8	7	6
	<i>All poultry</i>	5	5	5	4	5	5
Average path length	<i>Chickens</i>	2.11	2.11	2.12	2.06	2.11	2.08
	<i>Bantams</i>	2.72	2.86	2.92	2.82	2.93	3.05
	<i>Ducks</i>	2.89	2.89	3.01	3.06	2.82	2.90
	<i>All poultry</i>	2.00	2.03	2.04	2.00	2.05	2.03
Clustering coefficient	<i>Chickens</i>	0.49	0.47	0.46	0.49	0.48	0.46
	<i>Bantams</i>	0.30	0.31	0.29	0.27	0.29	0.26
	<i>Ducks</i>	0.27	0.27	0.30	0.28	0.30	0.30
	<i>All poultry</i>	0.52	0.51	0.49	0.51	0.51	0.49
Number nodes in GSCC	<i>Chickens</i>	119	119	119	114	114	111
	<i>Bantams</i>	93	102	91	84	80	74
	<i>Ducks</i>	84	84	90	88	77	83
	<i>All poultry</i>	121	123	122	118	119	115

Table C2 continues next page

Table C2 continued

Number nodes in GWCC	<i>Chickens</i>	130	131	129	130	133	134
	<i>Bantams</i>	122	119	122	115	115	117
	<i>Ducks</i>	119	119	115	116	121	112
	<i>All poultry</i>	130	131	130	130	133	134
Reciprocity	<i>Chickens</i>	0.47	0.45	0.50	0.48	0.46	0.43
	<i>Bantams</i>	0.34	0.31	0.30	0.30	0.26	0.25
	<i>Ducks</i>	0.27	0.27	0.28	0.31	0.34	0.28
	<i>All poultry</i>	0.50	0.49	0.52	0.51	0.49	0.46
Fragmentation	<i>Chickens</i>	0.16	0.18	0.15	0.23	0.27	0.31
	<i>Bantams</i>	0.42	0.27	0.45	0.47	0.52	0.60
	<i>Ducks</i>	0.50	0.50	0.39	0.43	0.60	0.45
	<i>All poultry</i>	0.13	0.12	0.12	0.18	0.20	0.26

Table C3. Summary data from a Susceptible-Infectious (SI) network stochastic simulation modelling poultry trades between 134 suburbs in New Zealand through the online auction website TradeMe*. In total 10,000 permutations were performed with the infection seeded randomly in a single suburb at the beginning of each permutation. The simulation was stopped either when all the suburb nodes had been infected or when a maximum of seven years (2,555 days) had elapsed. Suburbs are listed in alphabetical order.

Suburb	Total runs infected (%)	Mean number of days until infected (min-max)	Total runs infected within first 14 days (%)	Total runs infected within first 30 days (%)
Akaroa	9652 (96.52)	162.33 (0-2529)	1040 (10.78)	3039 (31.49)
Alexandra	9652 (96.52)	151.91 (0-2512)	1347 (13.96)	3833 (39.72)
Amberley	9652 (96.52)	144.15 (0-2509)	2560 (26.53)	5079 (52.63)
Ashburton	9652 (96.52)	142.09 (0-2510)	2947 (30.54)	5276 (54.67)
Auckland City	9652 (96.52)	138.71 (0-2511)	3819 (39.57)	5549 (57.50)
Balclutha	9652 (96.52)	148.20 (0-2518)	1877 (19.45)	4454 (46.15)
Blenheim	9652 (96.52)	144.90 (0-2513)	2314 (23.98)	4962 (51.41)
Bluff	9652 (96.52)	192.22 (0-2529)	437 (4.53)	1580 (16.37)
Bulls	9652 (96.52)	156.14 (0-2552)	1150 (11.92)	3374 (34.96)
Cambridge	9652 (96.52)	142.65 (0-2509)	2927 (30.33)	5173 (53.60)
Carterton	9652 (96.52)	151.92 (0-2513)	1706 (17.68)	4159 (43.09)
Chatham Islands	9708 (97.08)	352.11 (0-2542)	229 (2.36)	601 (6.20)
Cheviot	9652 (96.52)	159.10 (0-2543)	1030 (10.68)	3015 (31.24)
Christchurch City	9652 (96.52)	139.64 (0-2508)	3531 (36.59)	5485 (56.83)
Coromandel	9652 (96.52)	149.51 (0-2511)	1726 (17.89)	4213 (43.65)
Cromwell	9652 (96.52)	157.95 (0-2536)	1010 (10.47)	3079 (31.91)
Dannevirke	9652 (96.52)	152.68 (0-2547)	1589 (16.47)	4024 (41.70)
Darfield	9652 (96.52)	142.59 (0-2507)	2858 (29.62)	5243 (54.33)
Dargaville	9652 (96.52)	147.04 (0-2517)	2084 (21.60)	4661 (48.30)
Dunedin	9652 (96.52)	141.48 (0-2518)	3069 (31.80)	5348 (55.41)
Edendale	9652 (96.52)	153.99 (0-2516)	1130 (11.71)	3538 (36.66)
Fairlie	9652 (96.52)	157.32 (0-2529)	993 (10.29)	3225 (33.42)
Featherston	9652 (96.52)	153.70 (0-2528)	1361 (14.11)	3938 (40.80)
Feilding	9652 (96.52)	144.77 (0-2514)	2619 (27.14)	5043 (52.25)
Franklin	9652 (96.52)	140.03 (0-2511)	3587 (37.17)	5463 (56.60)
Geraldine	9652 (96.52)	154.75 (0-2514)	1405 (14.56)	3753 (38.89)
Gisborne	9652 (96.52)	145.81 (0-2521)	2036 (21.10)	4598 (47.64)

Table C3 continues next page

Table C3 continued

Golden Bay	9651 (96.51)	159.53 (0-2516)	829 (8.59)	2751 (28.51)
Gore	9652 (96.52)	147.00 (0-2510)	1989 (20.61)	4641 (48.09)
Great Barrier Island	9638 (96.38)	259.82 (0-2552)	275 (2.86)	851 (8.83)
Greymouth	9652 (96.52)	147.67 (0-2513)	2099 (21.75)	4600 (47.66)
Greytown	9652 (96.52)	162.25 (0-2545)	911 (9.44)	2981 (30.89)
Hamilton	9652 (96.52)	139.41 (0-2512)	3694 (38.28)	5473 (56.71)
Hanmer Springs	9652 (96.52)	157.03 (0-2514)	1287 (13.34)	3423 (35.47)
Hastings	9652 (96.52)	144.12 (0-2522)	2730 (28.29)	5066 (52.49)
Hawera	9652 (96.52)	144.14 (0-2514)	2671 (27.68)	5067 (52.50)
Helensville	9652 (96.52)	142.22 (0-2508)	3061 (31.72)	5256 (54.46)
Hibiscus Coast	9652 (96.52)	142.93 (0-2511)	2951 (30.58)	5200 (53.88)
Hokitika	9652 (96.52)	150.28 (0-2513)	1450 (15.03)	3944 (40.87)
Huntly	9652 (96.52)	142.94 (0-2515)	2892 (29.97)	5189 (53.77)
Invercargill	9652 (96.52)	144.31 (0-2510)	2437 (25.25)	5010 (51.91)
Kaiapoi	9652 (96.52)	144.59 (0-2511)	2500 (25.91)	4975 (51.55)
Kaikohe	9652 (96.52)	151.55 (0-2509)	1645 (17.05)	4029 (41.75)
Kaikoura	9652 (96.52)	156.30 (0-2519)	1151 (11.93)	3374 (34.96)
Kaitaia	9652 (96.52)	150.11 (0-2519)	1666 (17.27)	4281 (44.36)
Kapiti	9652 (96.52)	142.26 (0-2516)	3048 (31.58)	5303 (54.95)
Katikati	9652 (96.52)	146.36 (0-2521)	2198 (22.78)	4815 (49.89)
Kawakawa	9649 (96.49)	174.07 (0-2543)	735 (7.62)	2519 (26.11)
Kerikeri	9652 (96.52)	145.92 (0-2521)	2402 (24.89)	4831 (50.06)
Kurow	9652 (96.52)	170.74 (0-2525)	756 (7.84)	2567 (26.60)
Lawrence	9651 (96.51)	179.98 (0-2522)	523 (5.42)	1751 (18.15)
Levin	9652 (96.52)	145.07 (0-2511)	2448 (25.37)	4927 (51.05)
Lower Hutt City	9652 (96.52)	141.60 (0-2516)	3284 (34.03)	5369 (55.63)
Lumsden	9652 (96.52)	158.96 (0-2511)	948 (9.83)	2987 (30.95)
Manawatu	9652 (96.52)	148.14 (0-2518)	1911 (19.8)	4571 (47.36)
Manukau City	9652 (96.52)	138.83 (0-2508)	3827 (39.65)	5535 (57.35)
Marlborough Sounds	9652 (96.52)	158.97 (0-2542)	920 (9.54)	3098 (32.10)
Martinborough	9650 (96.50)	161.28 (0-2536)	777 (8.06)	2908 (30.14)
Marton	9652 (96.52)	152.69 (0-2533)	1465 (15.18)	4042 (41.88)
Masterton	9652 (96.52)	145.12 (0-2521)	2385 (24.71)	4997 (51.78)
Matamata	9652 (96.52)	144.65 (0-2511)	2523 (26.14)	4989 (51.69)
Maungaturoto	9652 (96.52)	150.45 (0-2513)	1635 (16.94)	4248 (44.02)
Milton	9652 (96.52)	155.91 (0-2519)	1160 (12.02)	3483 (36.09)
Mokau	9709 (97.09)	354.21 (0-2544)	161 (1.66)	497 (5.12)
Morrinsville	9652 (96.52)	142.31 (0-2518)	3053 (31.64)	5225 (54.14)

Table C3 continues next page

Table C3 continued

Motueka	9652 (96.52)	150.31 (0-2519)	1485 (15.39)	4080 (42.28)
Mt Cook	9722 (97.22)	245.27 (0-2526)	278 (2.86)	856 (8.81)
Mt. Maunganui	9652 (96.52)	150.15 (0-2516)	1553 (16.09)	4152 (43.02)
Murchison	9649 (96.49)	236.91 (0-2549)	261 (2.71)	843 (8.74)
Napier	9652 (96.52)	144.00 (0-2524)	2772 (28.72)	5084 (52.68)
Nelson	9652 (96.52)	146.13 (0-2515)	2298 (23.81)	4720 (48.91)
New Plymouth	9652 (96.52)	142.58 (0-2508)	3028 (31.38)	5220 (54.09)
North Shore	9652 (96.52)	139.83 (0-2511)	3621 (37.52)	5467 (56.65)
Oamaru	9652 (96.52)	146.65 (0-2516)	2115 (21.92)	4618 (47.85)
Ohakune	9652 (96.52)	163.21 (0-2517)	1002 (10.39)	2963 (30.70)
Opotiki	9652 (96.52)	159.48 (0-2524)	1032 (10.70)	3011 (31.20)
Opunake	9652 (96.52)	154.59 (0-2520)	1164 (12.06)	3657 (37.89)
Otautau	9652 (96.52)	153.43 (0-2520)	1223 (12.68)	3675 (38.08)
Otorohanga	9652 (96.52)	146.65 (0-2524)	2186 (22.65)	4758 (49.30)
Paeroa	9651 (96.51)	153.49 (0-2526)	1473 (15.27)	3879 (40.20)
Pahiatua	9652 (96.52)	154.01 (0-2535)	1226 (12.71)	3880 (40.20)
Paihia	9652 (96.52)	167.27 (0-2536)	846 (8.77)	2668 (27.65)
Palmerston	9652 (96.52)	156.07 (0-2510)	1078 (11.17)	3358 (34.80)
Palmerston North	9652 (96.52)	141.78 (0-2512)	3302 (34.22)	5383 (55.78)
Papakura City	9652 (96.52)	141.02 (0-2510)	3337 (34.58)	5381 (55.76)
Picton	9640 (96.4)	351.18 (0-2545)	186 (1.93)	478 (4.96)
Porirua	9652 (96.52)	143.43 (0-2514)	2922 (30.28)	5221 (54.10)
Queenstown	9652 (96.52)	153.65 (0-2539)	1390 (14.41)	3824 (39.62)
Raglan	9652 (96.52)	154.20 (0-2514)	1578 (16.35)	4072 (42.19)
Ranfurly	9652 (96.52)	164.26 (0-2516)	870 (9.02)	2638 (27.34)
Rangiora	9652 (96.52)	141.14 (0-2510)	3216 (33.32)	5387 (55.82)
Riverton	9652 (96.52)	170.34 (0-2525)	781 (8.10)	2532 (26.24)
Rotorua	9652 (96.52)	141.81 (0-2510)	3159 (32.73)	5289 (54.80)
Roxburgh	9650 (96.5)	187.58 (0-2523)	480 (4.98)	1601 (16.60)
Ruatoria	9695 (96.95)	315.98 (0-2538)	187 (1.93)	525 (5.42)
Stewart Island	9652 (96.52)	224.43 (0-2528)	322 (3.34)	1152 (11.94)
Stratford	9652 (96.52)	146.23 (0-2514)	2421 (25.09)	4826 (50.00)
Taihape	9652 (96.52)	163.75 (0-2536)	756 (7.84)	2708 (28.06)
Taumarunui	9652 (96.52)	151.88 (0-2518)	1468 (15.21)	3876 (40.16)
Taupo	9652 (96.52)	144.94 (0-2524)	2513 (26.04)	5013 (51.94)
Tauranga	9652 (96.52)	141.27 (0-2511)	3305 (34.25)	5347 (55.40)
Te Anau	9652 (96.52)	165.17 (0-2530)	705 (7.31)	2530 (26.22)
Te Awamutu	9652 (96.52)	142.49 (0-2508)	3036 (31.46)	5248 (54.38)

Table C3 continues next page

Table C3 continued

Te Kuiti	9652 (96.52)	147.60 (0-2524)	1999 (20.72)	4516 (46.79)
Te Puke	9652 (96.52)	144.98 (0-2521)	2429 (25.17)	4974 (51.54)
Thames	9652 (96.52)	146.15 (0-2538)	2312 (23.96)	4786 (49.59)
Timaru	9652 (96.52)	144.70 (0-2508)	2398 (24.85)	5008 (51.89)
Tokanui	9652 (96.52)	190.17 (0-2519)	354 (3.67)	1252 (12.98)
Tokoroa/Putaruru	9652 (96.52)	144.31 (0-2512)	2531 (26.23)	5039 (52.21)
Turangi	9653 (96.53)	193.69 (0-2547)	397 (4.12)	1225 (12.70)
Twizel	9652 (96.52)	172.59 (0-2548)	501 (5.20)	1804 (18.70)
Upper Hutt City	9652 (96.52)	143.66 (0-2516)	2844 (29.47)	5174 (53.61)
Waiheke Island	9652 (96.52)	156.08 (0-2523)	1183 (12.26)	3272 (33.9)
Waihi	9652 (96.52)	153.66 (0-2536)	1496 (15.50)	3833 (39.72)
Waihi Beach	9652 (96.52)	205.93 (0-2533)	340 (3.53)	1196 (12.4)
Waimate	9652 (96.52)	151.32 (0-2514)	1525 (15.80)	3988 (41.32)
Waiouru	9655 (96.55)	277.14 (0-2542)	177 (1.84)	582 (6.03)
Waipukurau	9652 (96.52)	146.96 (0-2530)	2177 (22.56)	4718 (48.89)
Wairoa	9652 (96.52)	158.33 (0-2520)	1058 (10.97)	3218 (33.35)
Waitakere City	9652 (96.52)	139.50 (0-2510)	3702 (38.36)	5488 (56.86)
Wanaka	9652 (96.52)	154.05 (0-2524)	1239 (12.84)	3577 (37.06)
Wanganui	9652 (96.52)	160.31 (0-2527)	2382 (24.68)	4570 (47.35)
Warkworth	9652 (96.52)	145.18 (0-2515)	2439 (25.27)	4937 (51.16)
Wellington City	9652 (96.52)	140.21 (0-2514)	3588 (37.18)	5520 (57.20)
Wellsford	9652 (96.52)	146.09 (0-2512)	2249 (23.31)	4795 (49.68)
Westport	9651 (96.51)	157.62 (0-2544)	1329 (13.78)	3584 (37.14)
Whakatane	9653 (96.53)	143.36 (0-2552)	2814 (29.16)	5166 (53.52)
Whangamata	9651 (96.51)	199.87 (0-2529)	424 (4.40)	1411 (14.63)
Whanganui	9184 (91.84)	1092.09 (0-2552)	354 (3.86)	651 (7.09)
Whangarei	9652 (96.52)	142.01 (0-2508)	3127 (32.40)	5293 (54.84)
Whitianga	9521 (95.21)	547.96 (0-2555)	307 (3.23)	977 (10.27)
Winton	9652 (96.52)	150.74 (0-2510)	1461 (15.14)	4110 (42.59)
Woodville	9652 (96.52)	170.27 (0-2518)	661 (6.85)	2282 (23.65)

Table C4. List of bird species known to migrate to, from or within New Zealand excluding albatrosses, petrels, skuas and gannets. Data retrieved from the online encyclopaedia of New Zealand birds in April 2019 (<http://nzbirdsonline.org.nz>) and crossed checked with bird sightings reported to eBird between 01st January 2012 and 31st December 2018. All 32 species listed had been reported in eBird.

Common name	Scientific name
Wrybill	<i>Anarhynchus frontalis</i>
Ruddy turnstone	<i>Arenaria interpres</i>
Sharp-tailed sandpiper	<i>Calidris acuminata</i>
Sanderling	<i>Calidris alba</i>
Lesser knot	<i>Calidris canutus</i>
Curlew sandpiper	<i>Calidris ferruginea</i>
Pectoral sandpiper	<i>Calidris melanotos</i>
Red-necked stint	<i>Calidris ruficollis</i>
Banded dotterel	<i>Charadrius bicinctus</i>
Large sand dotterel	<i>Charadrius leschenaultii</i>
Lesser sand plover	<i>Charadrius mongolus</i>
White-winged black tern	<i>Chlidonias leucopterus</i>
Black-billed gull	<i>Chroicocephalus bulleri</i>
Shining cuckoo	<i>Chrysococcyx lucidus</i>
South Island pied oystercatcher	<i>Haematopus finschi</i>
Pied stilt	<i>Himantopus leucocephalus</i>
Eastern bar-tailed godwit	<i>Limosa lapponica</i>
Black-tailed godwit	<i>Limosa limosa</i>
Eastern curlew	<i>Numenius madagascariensis</i>
Whimbrel	<i>Numenius phaeopus</i>
Pacific golden plover	<i>Pluvialis fulva</i>
Grey plover	<i>Pluvialis squatarola</i>
Common tern	<i>Sterna hirundo</i>
Arctic tern	<i>Sterna paradisaea</i>
White-fronted tern	<i>Sterna striata</i>
Little tern	<i>Sternula albifrons</i>
Grey-tailed tattler	<i>Tringa brevipes</i>
Wandering tattler	<i>Tringa incana</i>
Greenshank	<i>Tringa nebularia</i>
Marsh sandpiper	<i>Tringa stagnatilis</i>
Long-tailed cuckoo	<i>Urodynamis taitensis</i>
Terek sandpiper	<i>Xenus cinereus</i>

Table C5. List of bird species belonging to Anatidae; a family of water birds that includes ducks, geese, and swans, found in New Zealand. Data retrieved from the online encyclopaedia of New Zealand birds in November 2019 (<http://nzbirdsonline.org.nz>) and crossed checked with bird sightings reported to eBird between 01st January 2012 and 31st December 2018. Only 91.7% (22/24) of the species listed had been reported in eBird excluding the northern pintail (*Anas acuta*) and the pink-eared duck (*Malacorhynchus membranaceus*).

Common name	Scientific name
Northern pintail	<i>Anas acuta</i>
Auckland Island teal	<i>Anas aucklandica</i>
Chestnut teal	<i>Anas castanea</i>
Brown teal	<i>Anas chlorotis</i>
Grey teal	<i>Anas gracilis</i>
Campbell Island teal	<i>Anas nesiotis</i>
Mallard	<i>Anas platyrhynchos</i>
Grey duck	<i>Anas superciliosa</i>
Greylag goose	<i>Anser anser</i>
Australian white-eyed duck	<i>Aythya australis</i>
New Zealand scaup	<i>Aythya novaeseelandiae</i>
Canada goose	<i>Branta canadensis</i>
Muscovy duck	<i>Cairina moschata</i>
Cape Barren goose	<i>Cereopsis novaehollandiae</i>
Australian wood duck	<i>Chenonetta jubata</i>
Black swan	<i>Cygnus atratus</i>
Mute swan	<i>Cygnus olor</i>
Plumed whistling duck	<i>Dendrocygna eytoni</i>
Blue duck	<i>Hymenolaimus malacorhynchos</i>
Pink-eared duck	<i>Malacorhynchus membranaceus</i>
Northern shoveler	<i>Spatula clypeata</i>
Australasian shoveler	<i>Spatula rhynchotis</i>
Chestnut-breasted shelduck	<i>Tadorna tadornoides</i>
Paradise shelduck	<i>Tadorna variegata</i>

Figure C1. Annual trader contact networks for all poultry trades occurring through the online auction website TradeMe® in New Zealand from 2012 to 2017. To improve clarity in the visualization, only connections to and from nodes with a degree ≥ 100 are shown with node colour and size highlighting those nodes with a degree centrality measure ≥ 100 (orange: degree < 100 , red: degree ≥ 100). For each year the total number of traders and number of movements is shown in parenthesis in the form of; YYYY (number of traders: number of movements).

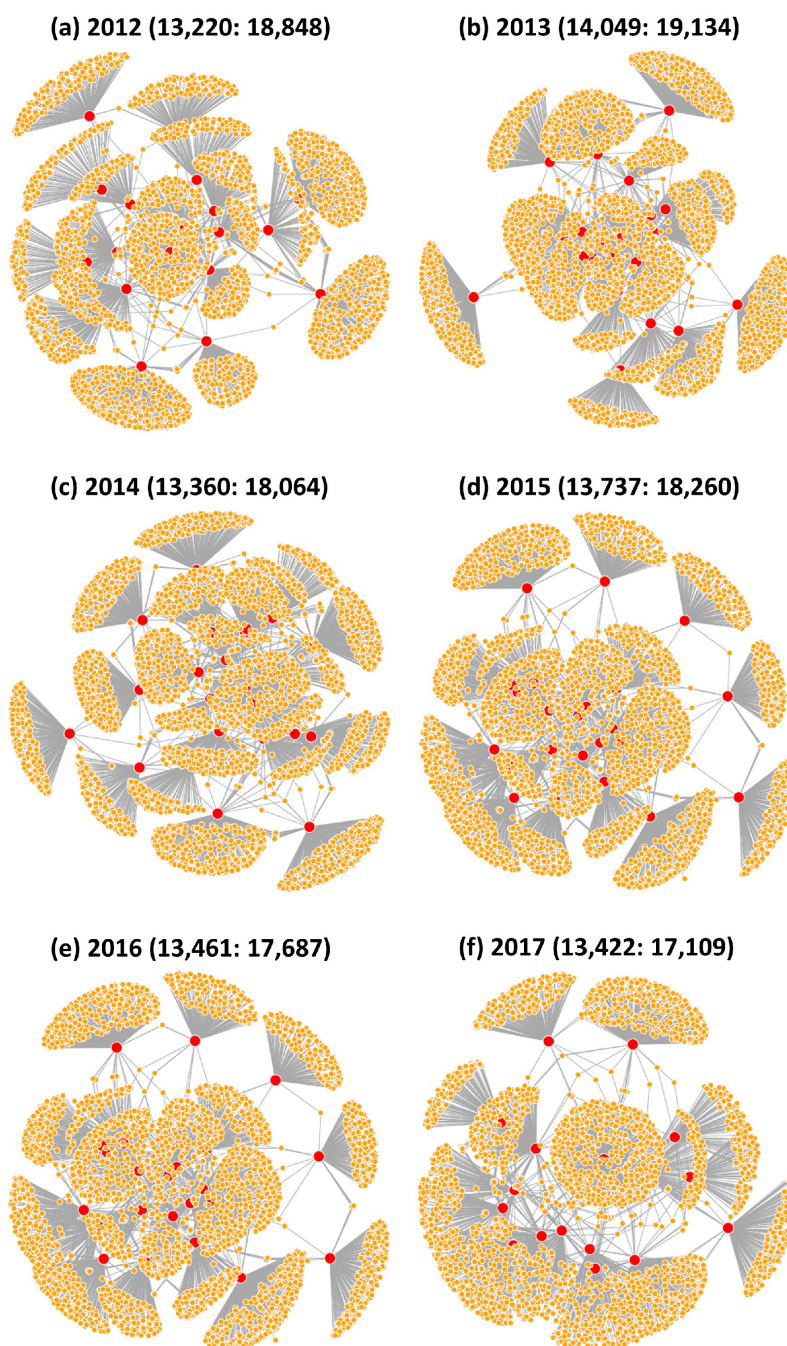
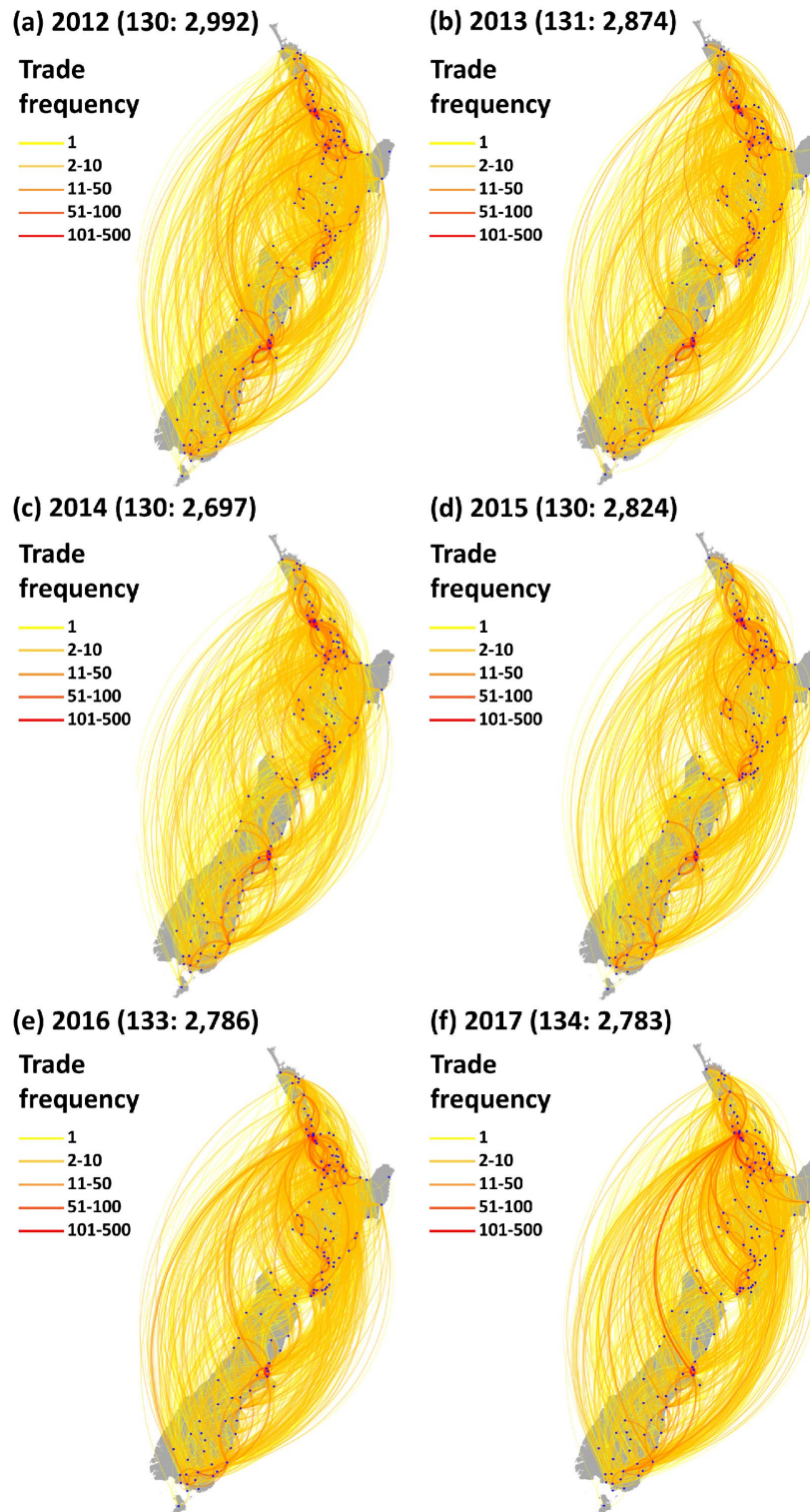


Figure C2. Annual spatial networks for all poultry trades occurring through the online auction website TradeMe® in New Zealand from 2012 to 2017. Edge colour indicates the frequency of trades (*i.e.*, the number of days between two consecutive trades going to and from the same two nodes). For each year the total number of suburbs and number of movements is shown in parenthesis in the form of; YYYY (number of suburbs: number of movements).



Supplementary Material
Chapter 5



Figure D2. (a) Network graph showing the four link communities identified and (b) the hierarchical cluster dendrogram used to define the cut-off point for community extraction (0.92). Network edges are coloured by the community they belong to whilst nodes are coloured according to the type of poultry enterprise (blue: poultry grower, red: poultry breeder) with size proportional to their node degree centrality measure.

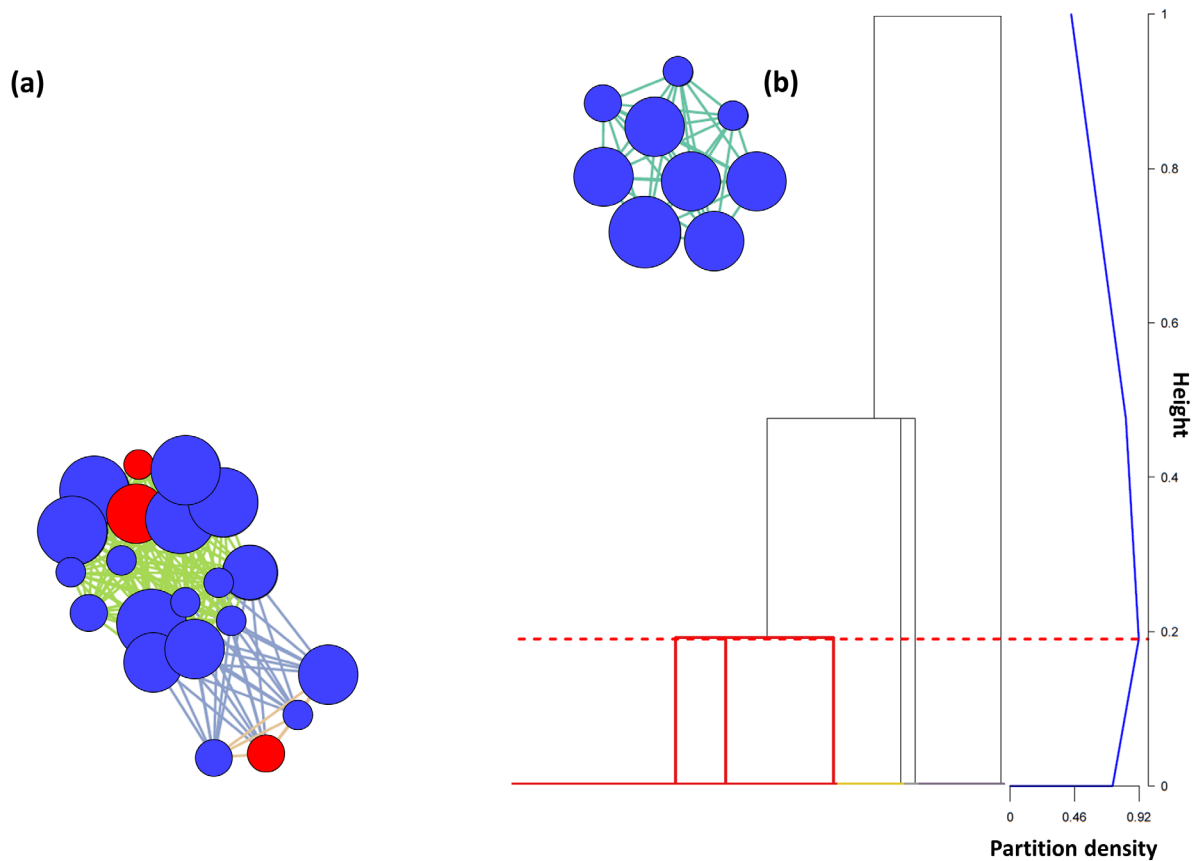


Figure D3. Imputed unimodal network graphs (left) and corresponding bimodal network graphs (right) showing (a, b) all on- and off-farm movements relating to feed, live birds, hatching eggs, waste or new and used litter between New Zealand commercial poultry farms ($n = 30$). Movements have been sub-divided so that the network graphs (c and d) show only movements relating to feed, (e and f) show only movements relating to live birds and hatching eggs and (g and h) show only movements relating to waste and litter. The node colour indicates the type of node (green: transporting company, blue: poultry grower, red: poultry breeder) whilst the node size is proportional to node degree centrality measure.

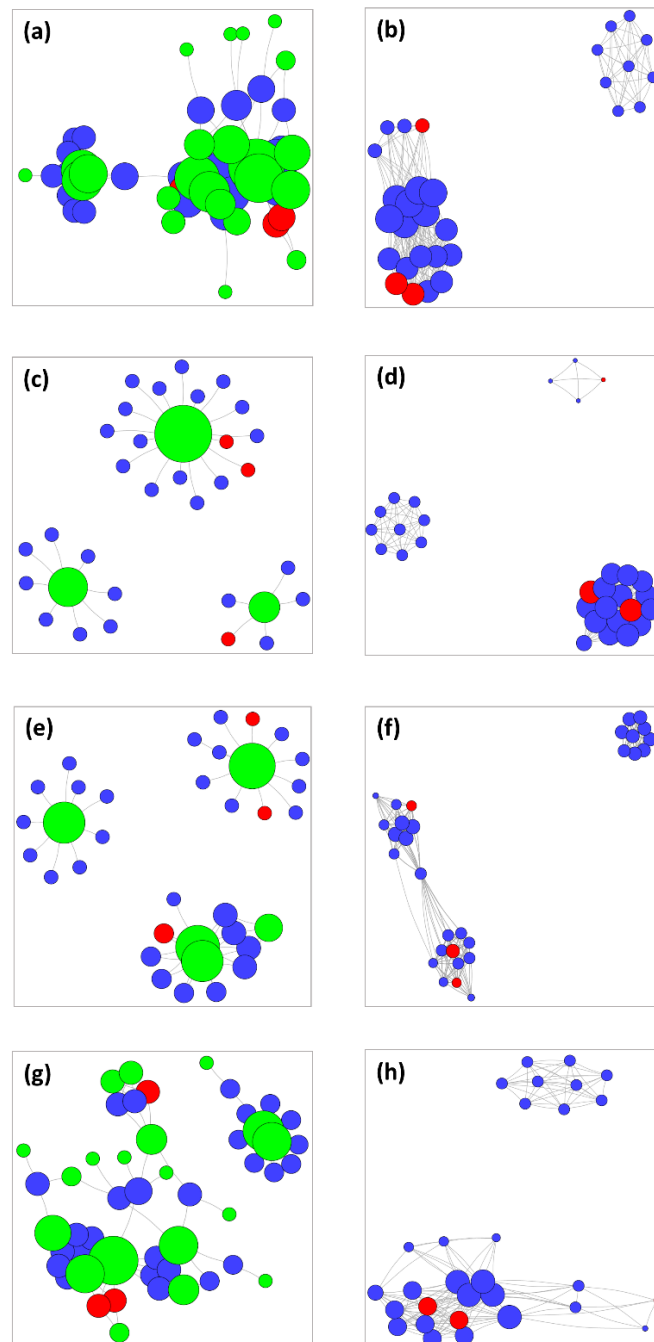


Table D1. PERMANOVA model comparing the allelic dissimilarity between 167 *C. jejuni* isolates by parent company (n = 3) and farm (nested within parent company) (n = 30) with *p*-values for each test obtained using 9999 unrestricted permutations. “SS” provides the sum of squares, “MS” the mean squares, and “*df*” the degrees of freedom for each test.

	<i>df</i>	SS	MS	Pseudo- <i>F</i>	<i>p</i> -value
Parent company	2	24308	12154	11.807	0.0001
Farm	27	26554	983.47	3.2949	0.0001
Residuals	137	40891	298.48		
Total	166	91753			

Table D2. Resemblance matrix considering the relationship between the allelic dissimilarity matrix and six additional matrix models indicating the network distance, geographical distance and shared parent companies between 167 *C. jejuni* isolates. The matrix was obtained using the Primer (v7.0.) 2STAGE routine using methods described in Anderson *et al.* (2017).

	Allelic dissimilarity	Euclidean distance	Feed network	Live bird network	Parent company	Road distance	Waste network
Allelic dissimilarity	-						
Euclidean distance	0.580	-					
Feed network	0.632	0.869	-				
Live bird network	0.623	0.895	0.941	-			
Parent company	0.438	0.597	0.762	0.744	-		
Road distance	0.584	0.988	0.869	0.905	0.597	-	
Waste network	0.614	0.890	0.940	0.973	0.723	0.899	-

Table D3. Resemblance matrix considering the relationship between the residual allelic dissimilarity matrix, after removing the effects of parent company and farms nested within parent company) and five additional matrix models indicating the network distance and geographical distance between 167 *C. jejuni* isolates. The matrix was obtained using the Primer (v7.0.) 2STAGE routine using methods described in Anderson et al (2017).

	Residual allelic dissimilarity	Euclidean distance	Feed network	Live bird network	Road distance	Waste network
Residual allelic dissimilarity	-					
Euclidean distance	-0.024	-				
Feed network	0.040	0.869	-			
Live bird network	0.001	0.895	0.941	-		
Road distance	-0.005	0.988	0.869	0.905	-	
Waste network	0.005	0.890	0.940	0.973	0.899	-

Table D4. PERMANOVA model comparing the allelic distance between 125 *C. jejuni* isolates by parent company (n = 3) and farm (nested within parent company) (n = 16) with *p*-values for each test obtained using 9999 unrestricted permutations. “SS” provides the sum of squares, “MS” the mean squares, and “*df*” the degrees of freedom for each test.

	<i>df</i>	SS	MS	Pseudo- <i>F</i>	<i>p</i> -value
Parent company	2	11531	11531	9.6467	0.0002
Farm	13	16931	1302.4	3.9195	0.0001
Residuals	109	36219	332.28		
Total	124	76211			

Table D5. Spearman's rank matrix correlation (ρ) between each model matrix and (i) the allelic dissimilarity matrix between 125 *C. jejuni* isolates and (ii) the residual allelic dissimilarity matrix after fitting the ANOVA factors of parent company ($n = 3$) and farm ($n = 16$) nested within parent company, with p -values obtained using 9999 permutations.

	(i) Unconstrained matrix		(ii) Residual matrix	
	ρ	p -value	ρ	p -value
Feed	0.665	0.0001	0.084	0.0263
Live birds	0.665	0.0001	0.076	0.0318
Waste	0.674	0.0001	0.108	0.0085
Road distance	0.609	0.0001	0.022	0.3162
Euclidean distance	0.603	0.0001	-0.017	0.6146
Parent company	0.573	0.0001	-	-

Table D6. Individual distance-based redundancy analysis (dbRDA) models to explain variation in allelic dissimilarities among 125 *C. jejuni* isolates in response to each of two ANOVA factors (parent company or farm nested in parent company) or sets of regression coordinates corresponding to geographic position (Euclidean distance or road distance) or the network model matrices of interest (feed, live birds or waste), with p -values for each of these separate marginal tests obtained using 9999 unrestricted permutations. “Prop” gives the proportion of the total variation explained whilst “ df ” gives the numerator (regression) and denominator (residual) degrees of freedom for the test. Models have been presented in order of decreasing R^2 values.

	Prop.	Pseudo- F	df	p -value
Parent Company	0.3026	26.47	3, 122	0.0001
Farm	0.5248	8.02	16, 109	0.0001
Euclidean distance	0.3539	33.42	3, 122	0.0001
Road distance	0.4128	28.35	4, 121	0.0001
Feed	0.3236	19.30	4, 121	0.0001
Live birds	0.2996	26.10	3, 122	0.0001
Waste	0.2898	24.89	3, 122	0.0001

Table D7. Distance-based redundancy analysis (dbRDA) to explain variation in allelic dissimilarities among 125 *C. jejuni* isolates in response to the factors and sets of regressors listed in Table S6, but here conditional tests were done in a sequential stepwise fashion under forward selection based on R^2 . Each test used 9999 permutations of residuals under a reduced model. “Prop” gives the proportion of additional variation explained by adding that set of variables to the model, “Cumul” tracks the cumulative explained variation with each added step, and “ df ” provides the regression and residual degrees of freedom.

Step		Prop.	Cumul.	df	Pseudo- F	p -value
1	+Farm	0.5248	0.5248	16, 109	8.024	0.0001
2	+Road	0.0286	0.5533	19, 106	2.260	0.0624
3	+Feed	0.0222	0.5755	22, 103	1.792	0.1267
4	+Live birds	0.0064	0.5818	24, 101	0.767	0.5288
5	+Euclidean	0.0041	0.5859	26, 99	0.490	0.6234
6	+Waste	0.0026	0.5886	28, 97	0.309	0.9057
7	+Parent company	<0.0001	0.5886	28, 97	-	-

Table D8. Top four dbRDA models obtained on the basis of the multivariate analogue to the Akaike Information Criterion (AIC) to explain variation in allelic dissimilarities among 125 *C. jejuni* isolates. Two ANOVA factors (parent company (n = 3) and farms (n = 16) nested in parent company) were included in all potential models *a priori*.

Model selections	R ²	No. Sets	AIC
Feed network, road distance, parent company and farm	0.5755	4	738.52
Road distance, parent company and farm	0.5533	3	738.88
Parent company and farm	0.5248	2	740.63
Feed network, live bird network, road distance, parent company and farm	0.5818	5	740.64

Table D9. PERMANOVA model comparing the p-distance between 167 *C. jejuni* isolates by parent company (n = 3) and farms nested within parent company (n = 30) with *p*-values for each test obtained using 9999 unrestricted permutations. “SS” provides the sum of squares, “MS” the mean squares, and “*df*” the degrees of freedom for each test.

	<i>df</i>	SS	MS	Pseudo- <i>F</i>	<i>p</i> -value
Parent company	2	33.159	0.0366	33.159	0.0001
Farm	27	0.0285	0.0011	3.4202	0.0001
Residuals	137	0.0423	0.0003		
Total	166	0.1440			

Table D10. Spearman's rank matrix correlation (ρ) between each model matrix and (i) the p-distance matrix between 167 *C. jejuni* isolates and (ii) the residual p-distance matrix after fitting the ANOVA factors of parent company (n = 3) and farm (n = 30) nested within parent company, with p-values obtained using 9999 permutations.

	(i) Unconstrained matrix		(ii) Residual matrix	
	ρ	p -value	ρ	p -value
Feed	0.663	0.0001	0.059	0.0061
Live birds	0.640	0.0001	0.006	0.2738
Waste	0.624	0.0001	-0.003	0.5766
Road distance	0.559	0.0001	0.011	0.3364
Euclidean distance	0.559	0.0001	0.013	0.3160
Parent company	0.499	0.0001	-	-

Table D11. Individual distance-based redundancy analysis (dbRDA) models to explain variation in the p-distance matrix among 167 *C. jejuni* isolates in response to each of two ANOVA factors (parent company or farm nested in parent company) or sets of regression coordinates corresponding to geographic position (Euclidean distance or road distance) or the network model matrices of interest (feed, live birds or waste), with *p*-values for each of these separate marginal tests obtained using 9999 unrestricted permutations. “Prop” gives the proportion of the total variation explained whilst “*df*” gives the numerator (regression) and denominator (residual) degrees of freedom for the test. Models have been presented in order of decreasing R^2 values.

	Prop.	Pseudo-<i>F</i>	<i>df</i>	<i>p</i>-value
Parent Company	0.5088	84.95	3, 164	0.0001
Farm	0.1978	1.269	28, 139	0.1251
Euclidean distance	0.5881	117.1	3, 164	0.0001
Road distance	0.5869	77.19	4, 163	0.0001
Feed	0.5966	80.36	4, 163	0.0001
Live birds	0.5353	94.46	3, 164	0.0001
Waste	0.5249	90.59	3, 164	0.0001

Table D12. Distance-based redundancy analysis (dbRDA) to explain variation in the p-distance matrix among 167 *C. jejuni* isolates in response to the factors and sets of regressors listed in Table S11, but here conditional tests were done in a sequential stepwise fashion under forward selection based on R^2 . Each test used 9999 permutations of residuals under a reduced model. “Prop” gives the proportion of additional variation explained by adding that set of variables to the model, “Cumul” tracks the cumulative explained variation with each added step, and “ df ” provides the regression and residual degrees of freedom.

Step		Prop.	Cumul.	df	Pseudo- F	p -value
1	+Feed	0.5966	0.5966	4, 163	80.36	0.0001
2	+Farm	0.1094	0.7060	31, 136	1.874	0.0062
3	+Road	0.0122	0.7182	34, 133	1.912	0.0724
4	+Parent company	0.0064	0.7246	36, 131	1.526	0.1740
5	+Live birds	0.0037	0.7283	38, 129	0.884	0.4970
6	+Euclidean	0.0034	0.7318	40, 127	0.816	0.4330
7	+Waste	0.0018	0.7336	42, 125	0.429	0.8435

Table D13. Top four dbRDA models obtained on the basis of the multivariate analogue to the Akaike Information Criterion (AIC) to explain variation in the p-distance matrix among 167 *C. jejuni* isolates. Two ANOVA factors (parent company (n = 3) and farms (n = 30) nested in parent company) were included in all potential models *a priori*.

Model selections	R^2	No. Sets	AIC
Road distance, parent company and farm	0.7202	3	-1325.0
Euclidean distance, road distance, parent company and farm	0.7239	4	-1323.3
Parent company and farm	0.7066	2	-1323.1
Live birds, road distance, parent company and farm	0.7237	4	-1323.1

Table D14. PERMANOVA model comparing the p-distance between 125 *C. jejuni* isolates by parent company (n = 3) and farms nested within parent company (n = 16) with *p*-values for each test obtained using 9999 unrestricted permutations. “SS” provides the sum of squares, “MS” the mean squares, and “*df*” the degrees of freedom for each test.

	<i>df</i>	SS	MS	Pseudo- <i>F</i>	<i>p</i> -value
Parent company	2	0.049461	0.02473	21.913	0.0001
Farm	13	0.015983	0.001229	3.8939	0.0001
Residuals	109	0.034415	0.000316		
Total	124	0.099859			

Table D15. Spearman's rank matrix correlation (ρ) between each model matrix and (i) p-distance matrix between 125 *C. jejuni* isolates and (ii) the p-distance matrix after fitting the ANOVA factors of parent company ($n = 3$) and farm ($n = 16$) nested within parent company, with p-values obtained using 9999 permutations.

	(i) Unconstrained matrix		(ii) Residual matrix	
	ρ	p -value	ρ	p -value
Feed	0.649	0.0001	-0.039	0.8238
Live birds	0.647	0.0001	-0.066	0.9719
Waste	0.630	0.0001	-0.084	0.9862
Road distance	0.579	0.0001	-0.065	0.9120
Euclidean distance	0.575	0.0001	-0.065	0.8982
Parent company	0.540	0.0001	-	-

Table D16. Individual distance-based redundancy analysis (dbRDA) models to explain variation in the p-distance matrix among 125 *C. jejuni* isolates in response to each of two ANOVA factors (parent company or farm nested in parent company) or sets of regression coordinates corresponding to geographic position (Euclidean distance or road distance) or the network model matrices of interest (feed, live birds or waste), with *p*-values for each of these separate marginal tests obtained using 9999 unrestricted permutations. “Prop” gives the proportion of the total variation explained whilst “*df*” gives the numerator (regression) and denominator (residual) degrees of freedom for the test. Models have been presented in order of decreasing R^2 values.

	Prop.	Pseudo-<i>F</i>	<i>df</i>	<i>p</i>-value
Parent Company	0.4953	59.87	3, 122	0.0001
Farm	0.6554	13.82	16, 109	0.0001
Euclidean distance	0.5466	73.54	3, 122	0.0001
Road distance	0.5445	48.21	4, 121	0.0001
Feed	0.5489	49.07	4, 121	0.0001
Live birds	0.5393	71.40	3, 122	0.0001
Waste	0.5264	67.79	3, 122	0.0001

Table D17. Distance-based redundancy analysis (dbRDA) to explain variation within the p-distance matrix among 125 *C. jejuni* isolates in response to the factors and sets of regressors listed in Table S6, but here conditional tests were done in a sequential stepwise fashion under forward selection based on R^2 . Each test used 9999 permutations of residuals under a reduced model. “Prop” gives the proportion of additional variation explained by adding that set of variables to the model, “Cumul” tracks the cumulative explained variation with each added step, and “ df ” provides the regression and residual degrees of freedom.

Step		Prop.	Cumul.	df	Pseudo- F	p -value
1	+Farm	0.6554	0.6554	16, 109	13.82	0.0001
2	+Road	0.0270	0.6824	19, 106	3.006	0.0136
3	+Feed	0.0156	0.6979	22, 103	1.768	0.1245
4	+Euclidean	0.0057	0.7036	24, 101	0.965	0.3933
5	+Live birds	0.0031	0.7067	26, 99	0.519	0.7823
6	+Waste	0.0021	0.7088	28, 97	0.348	0.8770

Table D18. Top four dbRDA models obtained on the basis of the multivariate analogue to the Akaike Information Criterion (AIC) to explain variation in the p-distance matrix among 125 *C. jejuni* isolates. Two ANOVA factors (parent company (n = 3) and farms (n = 16) nested in parent company) were included in all potential models *a priori*.

Model selections	R^2	No. Sets	AIC
Feed network, road distance, parent company and farm	0.6979	4	-997.18
Road distance, parent company and farm	0.6824	3	-996.90
Euclidean distance, Feed network, road distance, parent company and farm	0.7036	5	-995.55
Live bird network, road distance, parent company and farm	0.6884	4	-995.29

Supplementary Material
Chapter 7

Table E1. GenBank accession numbers and the corresponding BioSample IDs of the 103 *S. aureus* Sequence Type (ST)-1 whole-genome sequence assemblies.

GenBank accession	BioSample
GCF_000011265.1	SAMD00061104
GCF_000011525.1	SAMEA1705922
GCF_000149015.1	SAMN00001489
GCF_000164715.1	SAMN00139434
GCF_000248655.1	SAMN00627551
GCF_000249015.1	SAMN00627578
GCF_000336515.1	SAMN02471422
GCF_000582745.1	SAMEA3138947
GCF_000626855.1	SAMN00138219
GCF_000626955.1	SAMN00117451
GCF_000747275.1	SAMN02983094
GCF_000763455.1	SAMN02953024
GCF_001018645.2	SAMN03255442
GCF_001019125.2	SAMN03255482
GCF_001019535.2	SAMN03255481
GCF_001063885.1	SAMN03197360
GCF_001275825.1	SAMN03753597
GCF_001297365.1	SAMN03658597
GCF_001469015.1	SAMN04191397
GCF_001680935.1	SAMN05188391
GCF_001681155.1	SAMN05188398
GCF_001879545.1	SAMN04243988
GCF_002096985.1	SAMN06546698
GCF_002125325.1	SAMN06698047
GCF_002188415.1	SAMN07167826
GCF_002188425.1	SAMN07167827
GCF_002188455.1	SAMN07167825
GCF_002188495.1	SAMN07167824
GCF_002209325.1	SAMN05520967
GCF_002572965.1	SAMN07765389
GCF_002633825.1	SAMN03763946
GCF_002795285.1	SAMN05864218
GCF_002798125.1	SAMN07967050
GCF_002887375.1	SAMN04346839
GCF_002905535.1	SAMN04089983
GCF_003017755.1	SAMN08456651
GCF_003038455.1	SAMN08683528

Table E1 continues next page.

Table E1 continued.

GenBank accession	BioSample
GCF_003038575.1	SAMN08683521
GCF_003038595.1	SAMN08683520
GCF_003237255.1	SAMN08717736
GCF_003237355.1	SAMN08717728
GCF_003237515.1	SAMN08717710
GCF_003238125.1	SAMN08717725
GCF_003238315.1	SAMN08717692
GCF_003238375.1	SAMN08717682
GCF_003238485.1	SAMN08717665
GCF_003238595.1	SAMN08717654
GCF_003238605.1	SAMN08717649
GCF_003238735.1	SAMN08717638
GCF_003238925.1	SAMN08717621
GCF_003238965.1	SAMN08717610
GCF_003239385.1	SAMN08717708
GCF_003239715.1	SAMN08717630
GCF_003239855.1	SAMN08717679
GCF_003240035.1	SAMN08717662
GCF_003240175.1	SAMN08717643
GCF_003336555.1	SAMN05853514
GCF_003421225.1	SAMD00083578
GCF_003422085.1	SAMD00083621
GCF_003422945.1	SAMD00083664
GCF_003573835.1	SAMD00134209
GCF_003605265.1	SAMN09935541
GCF_900020485.1	SAMEA1464707
GCF_900022645.1	SAMEA1464324
GCF_900038315.1	SAMEA1464506
GCF_900039205.1	SAMEA1464531
GCF_900040355.1	SAMEA1464588
GCF_900041185.1	SAMEA1317313
GCF_900041785.1	SAMEA1469688
GCF_900046035.1	SAMEA1464549
GCF_900046215.1	SAMEA1317392
GCF_900081355.1	SAMEA2384256
GCF_900081975.1	SAMEA2384527
GCF_900083415.1	SAMEA2445646
GCF_900097745.1	SAMEA2298730

Table E1 continues next page.

Table E1 continued.

GenBank accession	BioSample
GCF_900097885.1	SAMEA2298711
GCF_900097905.1	SAMEA2298740
GCF_900098195.1	SAMEA1708910
GCF_900124805.1	SAMEA3448998
GCF_900125475.1	SAMEA3448847
GCF_900125785.1	SAMEA3449400
GCF_900126045.1	SAMEA3448920
GCF_900126415.1	SAMEA3448890
GCF_900126535.1	SAMEA3449068
GCF_900127665.1	SAMEA3448997
GCF_900127805.1	SAMEA3448837
GCF_900127825.1	SAMEA3448860
GCF_900128065.1	SAMEA3448950
GCF_900128095.1	SAMEA3448960
GCF_900149245.1	SAMEA3883070
GCF_900149255.1	SAMEA3883072
GCF_900149335.1	SAMEA4535272
GCF_900250415.1	SAMEA104473384
GCF_900250665.1	SAMEA104473405
GCF_900250795.1	SAMEA104473424
GCF_900250985.1	SAMEA104473434
GCF_900251105.1	SAMEA104473460
GCF_900251355.1	SAMEA104473472
GCF_900251445.1	SAMEA104473490
GCF_900251525.1	SAMEA104473497
GCF_900252035.1	SAMEA104473547
GCF_900252205.1	SAMEA104473563
GCF_900457625.1	SAMEA3491707

Table E2. Network analysis glossary of terms used to describe the network representing the movement of live animals between dairy herds in New Zealand. Note: all the definitions given have been interpreted in context of this study. Measures include the number of nodes, number of edges, network diameter, average path length, in-degree, out-degree, betweenness, network density, clustering coefficient, number of giant strongly connected components (GSCC) and giant weakly connected components (GWCC), reciprocity, fragmentation and, assortativity.

Network metric		Definition
Network size	<i>Number of nodes</i>	Total number of farms recorded in the LIC-MINDA database sending or receiving animals
	<i>Number of edges</i>	Total number of live animal movements to or from all the farms in the network
	<i>Network diameter</i>	The longest path between any two pair of farms in the network
	<i>Average path length</i>	The average shortest path between any pair of farms in the network averaged over all pairs of farms
Centrality measures	<i>In-degree</i>	Number of movements received by a farm
	<i>Out-degree</i>	Number of movements sent off-farm
	<i>Betweenness</i>	The frequency a farm is found on the shortest path between any other two pair of farms in the network
Cohesion measures	<i>Network density</i>	The proportion of all possible links between farms in the network that are present
	<i>Clustering coefficient</i>	The proportion of neighbouring farms in direct contact with a farm that are also connected to each other
	<i>GSCC</i>	The largest sub-network in which all farms are mutually accessible by following the direction of the movements in the network
	<i>GWCC</i>	The largest sub-network in which all farms are linked when disregarding the direction of the movements in the network

Table E2 continues next page.

Table E2 continued.

Network metric	Definition
<i>Reciprocity</i>	The likelihood that any pair of farms in the network both receive and send animals between each other
<i>Fragmentation</i>	The proportion of farm pairs for which a movement does not exist between them
<i>Assortativity</i>	A preference for a farm to attach to other farms that have similar degree centrality measures

Table E3. Prevalence of 14 resistance genes identified across 57 *S. aureus* isolates isolated from bovine milk in New Zealand. Genes are listed in ascending order

Resistance Gene(s)	Number of Isolate (%)
<i>ant(9)-Ia, dfrC, fusC, mecA, mecR1</i>	1 (1.75)
<i>erm(A), qacB</i>	2 (3.51)
<i>blaPC1, qacA</i>	4 (7.02)
<i>fosD</i>	6 (10.53)
<i>blaZ</i>	18 (31.58)
<i>blaI, blaR1</i>	19 (33.33)
<i>tet(38)</i>	57 (100.0)

Table E4. Prevalence of 76 virulence genes identified across 57 *S. aureus* isolates isolated from bovine milk in New Zealand. Genes are listed in ascending order with a total of 55 genes being present in 100% of isolates

Virulence Gene(s)	Number of Isolates (%)
<i>chp, sea, selk</i>	1 (1.75)
<i>sak, scn</i>	2 (3.51)
<i>tsst-1</i>	3 (5.26)
<i>sec, sell</i>	4 (7.02)
<i>seh</i>	35 (61.40)
<i>cna</i>	41 (71.93)
<i>sdrD</i>	42 (73.68)
<i>fnbB</i>	45 (78.95)
<i>cap8(H-K)</i>	49 (85.96)
<i>esaC, esxB</i>	50 (87.72)
<i>sdrE</i>	51 (89.47)
<i>lukF-PV, map</i>	56 (98.25)
<i>adsA, aur, cap8(A-G), cap8(L-P), clf(A,B,P) coa, ebp, esa(A-B), (essA-C) esxA, fnbA, geh, hlb, hld, hlg(A-C), hly/hla, hysA, ica(A-D,R), isd(A-G), lip, sbi, sdrC, spa, srtB, ssp(A-C), vWbp</i>	57 (100.0)

Table E5. Antimicrobial sensitivity profiles of 50 *S. aureus* isolates labelled according to the isolate ID which identifies the date the sample was collected (dd/mm/yyyy), the farm from which it was collected from (A-Q), and the animal ID number (####). Sensitivity was determined using a zone diffusion test following the procedures provided by the Clinical and Laboratory Standards Institute (CLSI) for penicillin (Pen), novobiocin (Nov), ceftiofur (Cef), tetracycline (Tet), ceftiofur (XNL), and oxacillin (OXA). Zone range for each antimicrobial has been presented in millimetres (mm) with isolates being declared sensitive (S), intermediate (I) or resistant (R), based on CLSI recommendations.

Isolate ID	Sample type	Diameter (mm) (Sensitivity)					
		Pen	Nov	Cef	Tet	XNL	OXA
19Jan2016-A-90	Subclinical	18 (R)	36 (S)	30 (S)	31 (S)	32 (S)	21 (S)
27Jan2016-B-849	Subclinical	44 (S)	32 (S)	28 (S)	30 (S)	30 (S)	22 (S)
27Jan2016-B-1019	Subclinical	45 (S)	32 (S)	28 (S)	29 (S)	33 (S)	28 (S)
20Jan2016-C-554	Subclinical	46 (S)	40 (S)	28 (S)	32 (S)	34 (S)	26 (S)
07Jan2016-B-1061	Clinical	46 (S)	35 (S)	28 (S)	26 (S)	32 (S)	26 (S)
06Jan2016-D-321	Clinical	40 (S)	40 (S)	30 (S)	32 (S)	30 (S)	24 (S)
05Jan2016-E-452	Subclinical	44 (S)	36 (S)	29 (S)	30 (S)	32 (S)	26 (S)
27Jan2016-B-1323	Subclinical	40 (S)	32 (S)	28 (S)	32 (S)	32 (S)	22 (S)
20Jan2016-C-86	Subclinical	40 (S)	34 (S)	28 (S)	26 (S)	30 (S)	24 (S)
05Jan2016-E-452	Subclinical	44 (S)	36 (S)	28 (S)	30 (S)	32 (S)	25 (S)
08Jan2016-F-20	Clinical	40 (S)	36 (S)	26 (S)	29 (S)	36 (S)	24 (S)
21Jan2016-F-49	Clinical	42 (S)	38 (S)	28 (S)	31 (S)	31 (S)	24 (S)
05Jan2016-E-452	Subclinical	50 (S)	38 (S)	29 (S)	26 (S)	32 (S)	28 (S)
30Nov2015-H-163	Clinical	26 (R)	35 (S)	28 (S)	34 (S)	35 (S)	23 (S)
06Dec2015-(I)-50	Clinical	38 (S)	35 (S)	28 (S)	28 (S)	35 (S)	26 (S)
05Jan2016-E-80	Subclinical	40 (S)	34 (S)	27 (S)	26 (S)	30 (S)	26 (S)
13Jan2016-J-35	Subclinical	38 (S)	36 (S)	27 (S)	30 (S)	32 (S)	24 (S)
05Jan2016-E-80	Subclinical	44 (S)	36 (S)	26 (S)	27 (S)	30 (S)	26 (S)
13Jan2016-J-76	Subclinical	38 (S)	35 (S)	28 (S)	31 (S)	31 (S)	24 (S)
05Jan2016-E-80	Subclinical	42 (S)	36 (S)	26 (S)	28 (S)	31 (S)	26 (S)
13Jan2016-J-126	Subclinical	40 (S)	36 (S)	28 (S)	29 (S)	32 (S)	24 (S)
19Nov2015-E-222	Clinical	21 (R)	36 (S)	29 (S)	31 (S)	36 (S)	24 (S)
14Jan2016-K-695	Subclinical	11 (R)	34 (S)	28 (S)	29 (S)	27 (S)	16 (S)
13Jan2016-J-397	Subclinical	40 (S)	35 (S)	26 (S)	29 (S)	30 (S)	24 (S)
14Jan2016-K-501	Subclinical	40 (S)	35 (S)	30 (S)	29 (S)	29 (S)	26 (S)
13Jan2016-J-397	Subclinical	46 (S)	37 (S)	27 (S)	30 (S)	30 (S)	24 (S)
05Jan2016-E-315	Subclinical	41 (S)	34 (S)	27 (S)	30 (S)	30 (S)	26 (S)
15Dec2015-E-405	Clinical	22 (R)	34 (S)	28 (S)	26 (S)	30 (S)	25 (S)
05Jan2016-E-438	Subclinical	44 (S)	36 (S)	28 (S)	28 (S)	30 (S)	25 (S)

Table E5 continues next page.

Table E5 continued.

Isolate ID	Sample type	Diameter (mm) (Sensitivity)					
		Pen	Nov	Cef	Tet	XNL	OXA
05Jan2016-E-340	<i>Clinical</i>	22 (R)	34 (S)	26 (S)	28 (S)	30 (S)	24 (S)
20Nov2015-L-555	<i>Clinical</i>	49 (S)	36 (S)	28 (S)	30 (S)	33 (S)	30 (S)
14Jan2016-K-585	<i>Subclinical</i>	44 (S)	37 (S)	27 (S)	32 (S)	34 (S)	30 (S)
30Nov2015-M-885	<i>Clinical</i>	21 (R)	37 (S)	28 (S)	30 (S)	37 (S)	22 (S)
10Dec2015-N-28	<i>Subclinical</i>	16 (R)	36 (S)	26 (S)	26 (S)	36 (S)	19 (S)
14Jan2016-K-695	<i>Subclinical</i>	11 (R)	34 (S)	28 (S)	29 (S)	27 (S)	16 (S)
10Dec2015-N-53	<i>Subclinical</i>	14 (R)	36 (S)	28 (S)	30 (S)	26 (S)	12 (I)
22Nov2015-H-223	<i>Clinical</i>	19 (R)	40 (S)	30 (S)	33 (S)	40 (S)	22 (S)
10Dec2015-N-79	<i>Subclinical</i>	18 (R)	38 (S)	29 (S)	30 (S)	32 (S)	20 (S)
06Dec2015-O-261	<i>Clinical</i>	40 (S)	34 (S)	26 (S)	30 (S)	34 (S)	28 (S)
10Dec2015-N-242	<i>Subclinical</i>	12 (R)	30 (S)	27 (S)	28 (S)	25 (S)	12 (I)
07Jan2016-N-348	<i>Clinical</i>	14 (R)	26 (S)	28 (S)	30 (S)	25 (S)	12 (I)
10Dec2015-N-365	<i>Subclinical</i>	15 (R)	31 (S)	27 (S)	30 (S)	30 (S)	19 (S)
14Jan2016-K-117	<i>Subclinical</i>	14 (R)	34 (S)	29 (S)	27 (S)	29 (S)	15 (S)
14Dec2015-P-109	<i>Clinical</i>	34 (S)	36 (S)	28 (S)	30 (S)	30 (S)	28 (S)
20Jan2016-C-470	<i>Subclinical</i>	44 (S)	34 (S)	26 (S)	28 (S)	30 (S)	26 (S)
14Jan2016-K-478	<i>Subclinical</i>	13 (R)	34 (S)	29 (S)	28 (S)	29 (S)	16 (S)
19Nov2015-Q-416	<i>Clinical</i>	50 (S)	36 (S)	28 (S)	30 (S)	36 (S)	31 (S)
20Jan2016-C-508	<i>Subclinical</i>	50 (S)	36 (S)	34 (S)	32 (S)	34 (S)	30 (S)
19Jan2016-A-58	<i>Subclinical</i>	19 (R)	35 (S)	28 (S)	32 (S)	31 (S)	22 (S)
27Jan2016-B-809	<i>Subclinical</i>	44 (S)	34 (S)	27 (S)	28 (S)	30 (S)	26 (S)

Table E6. Three PERMANOVA models (1, 2, and 3) comparing the uncorrected p -dissimilarity measure between 57 *S. aureus* isolates by either township ($n = 12$), farm ($n = 17$) or, network community ($n = 6$) with p -values for each test obtained using 9999 unrestricted permutations. “SS” provides the sum of squares, “MS” the mean squares, and “ df ” the degrees of freedom for each test.

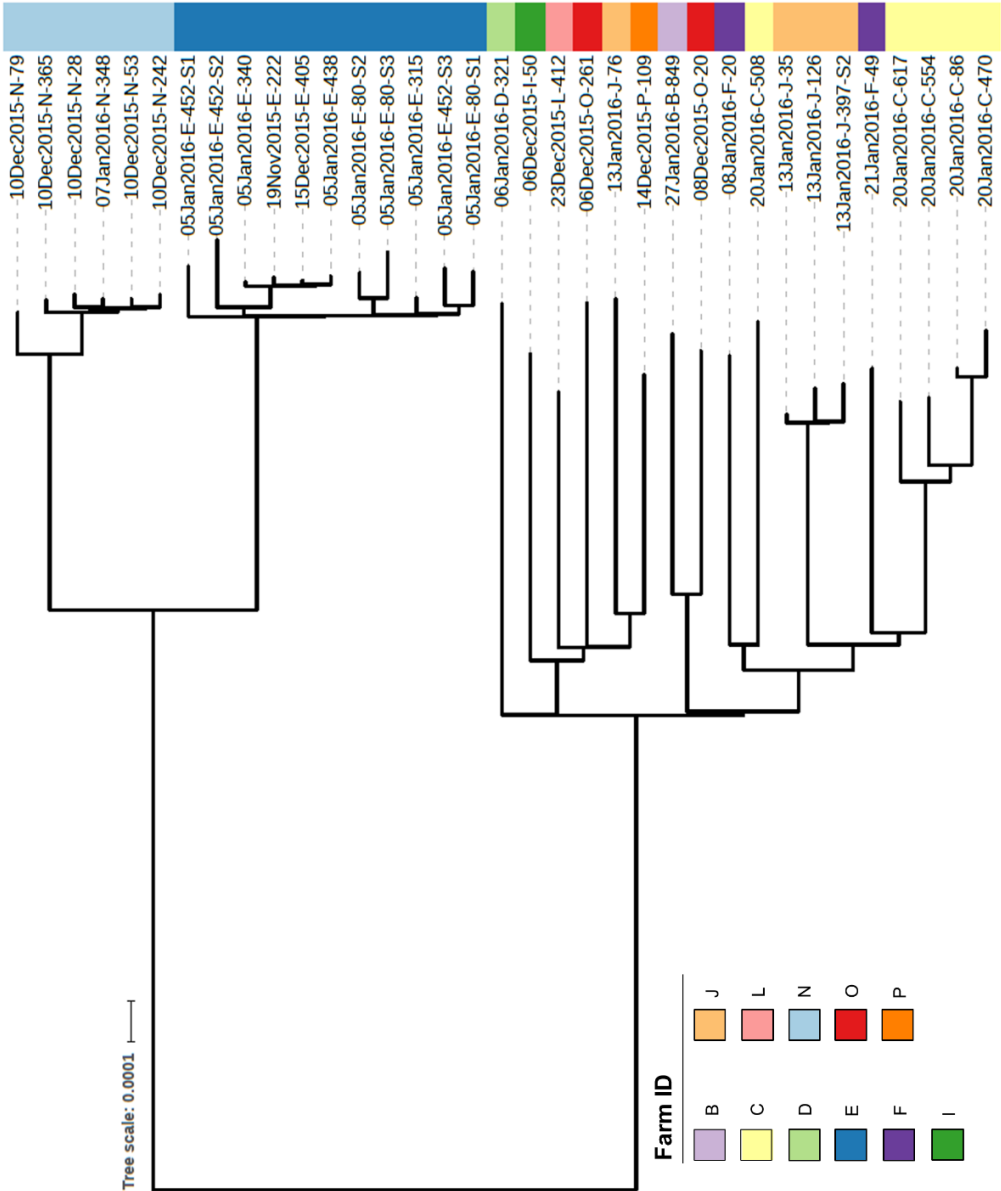
Model	Variables	df	SS	MS	Pseudo- F	p -value
1	Farm	16	1.15	0.07	2.06	0.0037
	Residuals	40	1.40	0.04		
2	Township	11	1.04	0.09	2.79	0.0004
	Residuals	45	1.52	0.03		
3	Community ^a	5	0.77	0.15	4.39	0.0002
	Residuals	51	1.79	0.04		

^a Communities identified using the network community algorithm based on greedy optimisation (Clauset *et al.* 2004)

Table E7. Pearson's Chi-squared test results between township (n = 12), farm (n = 17) and community (n = 6)

	Pearson's Chi-squared test (<i>df</i> , <i>p</i> -value)		
	Township	Farm	Community
Township	-	-	-
Farm	627.0 (176, 2.2×10^{-16})	-	-
Community	181.3 (55, 2.1×10^{-15})	285 (80, 2.2×10^{-16})	-

Figure E1. Maximum-likelihood phylogeny generated from core single nucleotide polymorphisms across 35 *S. aureus* Sequence Type (ST)-1 isolates. Isolate IDs identify the date the sample was collected (dd/mm/yyyy), the farm from which it was collected from (A-Q) as indicated in Figure 1, and the animal ID number (###)



Sample Type

- BM: bovine milk ●
- CC: canine clinical ★
- CN: canine nasal ■
- FC: feline clinical ★
- FP: feline perianal ▲
- HN: human nasal ■

Tree scale: [0.0000095]

Figure E3. Scatter plot showing the relationship between antimicrobial usage across 14 dairy herds in the Waikato region of New Zealand’s North Island and herd size. A linear model has been used to fit a regression line shown in blue with 95% confidence intervals shown in grey. Relationship has a Spearman’s rank correlation coefficient (SCC) equal to 0.27 ($p = 0.263$). AMU: antimicrobial usage, DD: daily dose

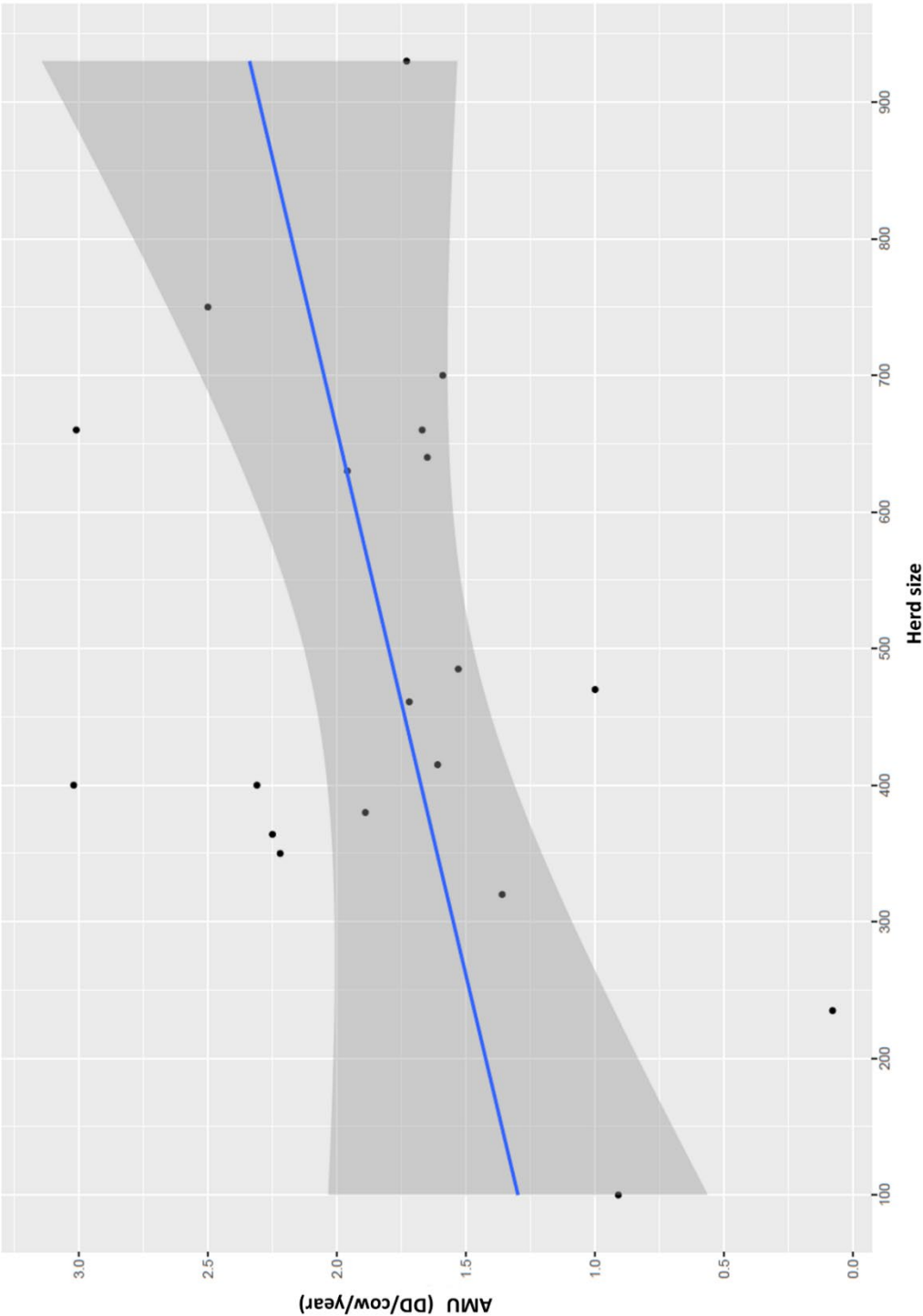


Figure E4. Scatter plot showing the relationship between antimicrobial usage, across 14 dairy herds in the Waikato region of New Zealand's North Island, and the number of resistant genes found to be present in isolates collected from each farm. The size of each point is proportional to the total number of genes present which the colour indicates which genes were present. AMU: antimicrobial usage, DD: daily dose

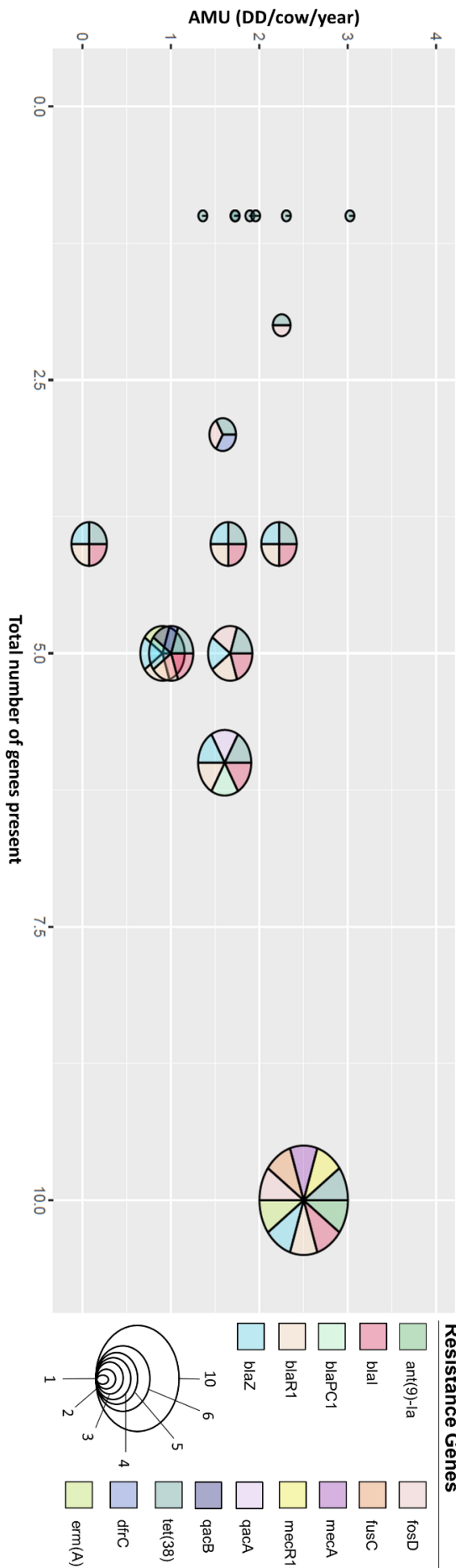


Figure E5. Multi-dimensional scaling (MDS) plots based on pairwise uncorrected *p*-measures between the core single nucleotide polymorphism profiles of 56 *S. aureus* isolates. The colour of the point indicates (a) the township the farm is located in ($n = 12$), (b) the farm ($n = 17$) from which the isolate was sample from and, (c) the community ($n = 6$) the farm is found in through the trade of live animals



Supplementary Material

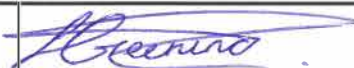
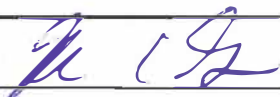
DRC16 Statement of Contribution forms



MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Sabrina Greening
Name/title of Primary Supervisor:	Associate Professor Carolyn Gates
Name of Research Output and full reference:	
Greening SS., Hidano A., French NP., Gates MC. Network Analysis in the Molecular Era: Untangling the Complex Relationships between Host Population Contact Structure, Disease Transmission Dynamics, and Pathogen Evolution (to be submitted)	
In which Chapter is the Manuscript /Published work:	Chapter 2
Please indicate:	
The percentage of the manuscript/Published Work that was contributed by the candidate:	80%
and	
Describe the contribution that the candidate has made to the Manuscript/Published Work:	
Sabrina conducted the majority of the literature review, wrote most of the manuscript, and is currently preparing the draft for submission.	
For manuscripts intended for publication please indicate target journal:	
Royal Society Interface	
Candidate's Signature:	
Date:	05/04/2020
Primary Supervisor's Signature:	
Date:	05/04/2020

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)


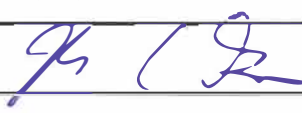
GRS Version 4– January 2019



MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Sabrina Greening
Name/title of Primary Supervisor:	Associate Professor Carolyn Gates
Name of Research Output and full reference:	
Greening SS., Mulqueen K., Rawdon TG., French NP., Gates MC. Estimating the level of disease risk and biosecurity on commercial poultry farms in New Zealand. <i>N. Z. Vet. J.</i> (In Press) 2020	
In which Chapter is the Manuscript /Published work:	Chapter 3 and Appendix B
Please indicate:	
The percentage of the manuscript/Published Work that was contributed by the candidate:	85%
and	
Describe the contribution that the candidate has made to the Manuscript/Published Work:	
Sabrina worked on data entry, wrote most of the scripts used to analyse the data, generated the tables and figures, wrote most of the manuscript, and edited the draft for publication.	
For manuscripts intended for publication please indicate target journal:	
New Zealand Veterinary Journal	
Candidate's Signature:	
Date:	05/04/2020
Primary Supervisor's Signature:	
Date:	05/04/2020

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)


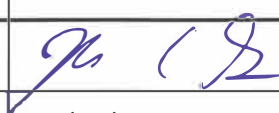
GRS Version 4– January 2019



MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Sabrina Greening	
Name/title of Primary Supervisor:	Associate Professor Carolyn Gates	
Name of Research Output and full reference:		
Greening SS., Mulqueen K., Rawdon TG., French NP., Gates MC. Characterising the disease risk from backyard poultry trade networks and migratory birds to New Zealand's commercial poultry industry. (to be submitted)		
In which Chapter is the Manuscript /Published work:	Chapter 4 and Appendix C	
Please indicate:		
The percentage of the manuscript/Published Work that was contributed by the candidate:	80%	
and		
Describe the contribution that the candidate has made to the Manuscript/Published Work:		
Sabrina generated the tables and figures, edited the draft manuscript, and is preparing it for submission.		
For manuscripts intended for publication please indicate target journal:		
Preventative Veterinary Medicine		
Candidate's Signature:		
Date:	05/04/2020	
Primary Supervisor's Signature:		
Date:	05/04/2020	

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)



GRS Version 4– January 2019



MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Sabrina Greening	
Name/title of Primary Supervisor:	Associate Professor Carolyn Gates	
Name of Research Output and full reference:		
Greening SS., Zhang J., Midwinter AC., Wilkinson DA., Fayaz A., Williamson DA., Anderson MJ., Gates MC., French NP. Transmission dynamics of <i>Campylobacter jejuni</i> (ST-6964) in New Zealand's commercial poultry network. (to be submitted)		
In which Chapter is the Manuscript /Published work:	Chapter 5 and Appendix D	
Please indicate:		
The percentage of the manuscript/Published Work that was contributed by the candidate:	80%	
and		
Describe the contribution that the candidate has made to the Manuscript/Published Work:		
Sabrina analysed the genomic data, wrote most of the scripts used in the analysis, generated the tables and figures, wrote most of the draft manuscript, and is preparing it for submission.		
For manuscripts intended for publication please indicate target journal:		
Epidemics		
Candidate's Signature:		
Date:	05/04/2020	
Primary Supervisor's Signature:		
Date:	05/04/2020	

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)

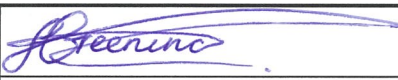
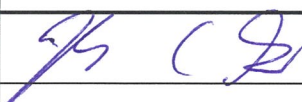
GRS Version 4– January 2019



MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Sabrina Greening
Name/title of Primary Supervisor:	Associate Professor Carolyn Gates
Name of Research Output and full reference:	
Greening SS., Zhang J., French NP., Gates MC. Assessing the impact of host contact network structure on the phylogeny of <i>Campylobacter jejuni</i> in commercial broiler production systems. (to be submitted)	
In which Chapter is the Manuscript /Published work:	Chapter 6
Please indicate:	
The percentage of the manuscript/Published Work that was contributed by the candidate:	85%
and	
Describe the contribution that the candidate has made to the Manuscript/Published Work:	
Sabrina wrote most of the scripts used to construct the model and analyse the data, generated the tables and figure, wrote the draft manuscript, and is preparing it for submission.	
For manuscripts intended for publication please indicate target journal:	
Veterinary Research	
Candidate's Signature:	
Date:	05/04/2020
Primary Supervisor's Signature:	
Date:	05/04/2020

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)

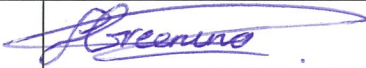
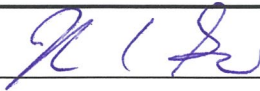
GRS Version 4– January 2019



MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Sabrina Greening		
Name/title of Primary Supervisor:	Associate Professor Carolyn Gates		
Name of Research Output and full reference:			
Greening SS., Zhang J., Midwinter AC., Wilkinson DA., McDougall S., Gates MC., French NP. Genetic relatedness of <i>Staphylococcus aureus</i> isolates obtained from New Zealand dairy cattle. (to be submitted)			
In which Chapter is the Manuscript /Published work:		Chapter 7 and Appendix E	
Please indicate:			
The percentage of the manuscript/Published Work that was contributed by the candidate:		80%	
and			
Describe the contribution that the candidate has made to the Manuscript/Published Work:			
Sabrina analysed the genomic data, wrote most of the scripts used in the analysis, generated the tables and figures, wrote most of the draft manuscript, and is preparing it for submission.			
For manuscripts intended for publication please indicate target journal:			
Scientific Reports			
Candidate's Signature:			
Date:	05/04/2020		
Primary Supervisor's Signature:			
Date:	05/04/2020		

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)

GRS Version 4– January 2019

[End of document]